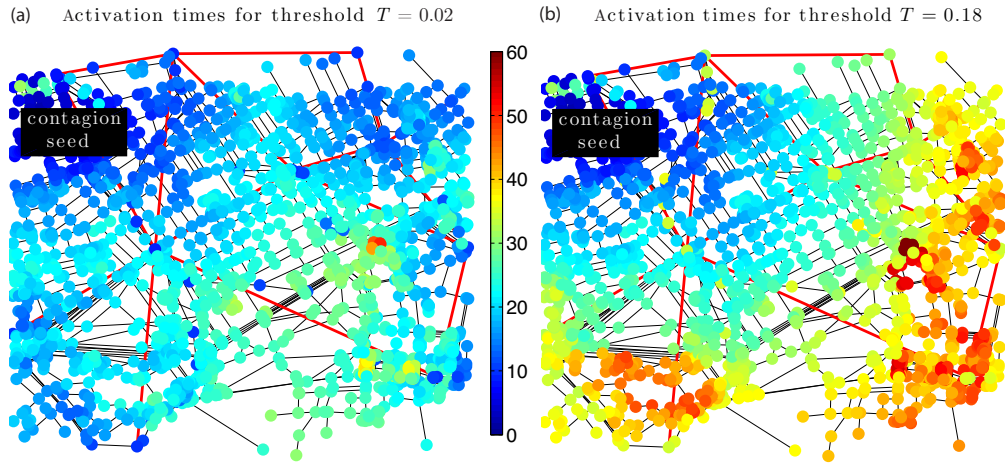
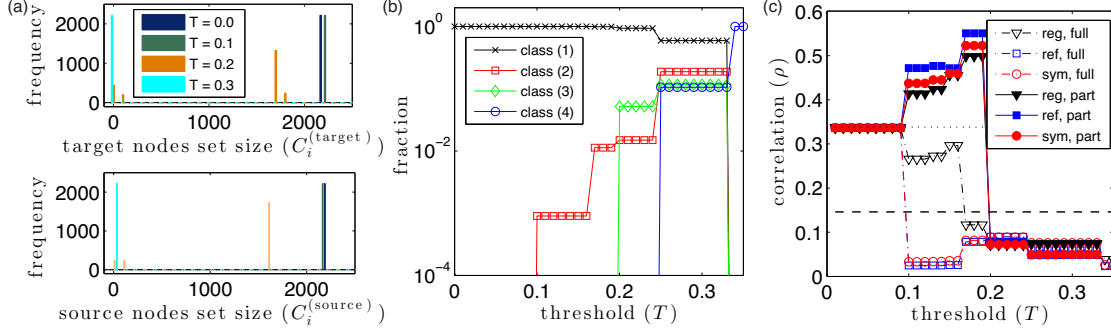


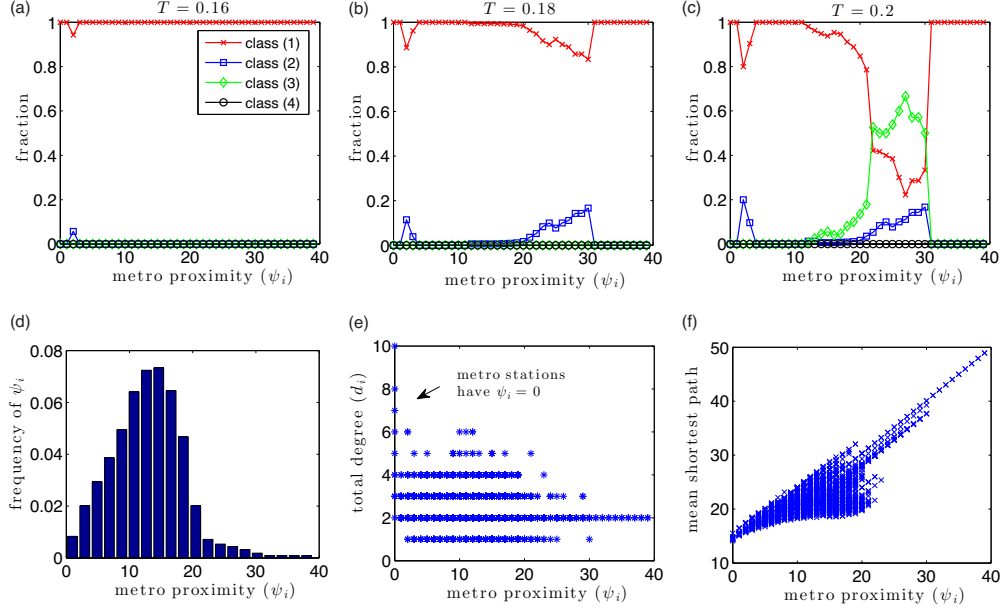
Supplementary Figure 1: London transit network with  $N = 2217$  nodes, 2854 geometric edges, and 15 non-geometric edges (which we have made publicly available, as we discussed in Sec. III A of the main manuscript). (a) The geometric edges (blue), which we take from Ref. [1], are roads between intersections; and the non-geometric edges (red), which we take from Ref. [2], give connections between metro stations. Some nodes ( $i \in \mathcal{P} \subset \mathcal{V}$ , where  $|\mathcal{P}| = 11$ ) correspond to both intersections and metro stations, whereas other nodes ( $i \in \mathcal{V} \setminus \mathcal{P}$ ) correspond only to intersections. Each node  $i \in \mathcal{V}$  has an intrinsic location  $\{\mathbf{w}^{(i)}\}$  based on its latitude and longitude. (b) Histogram of the frequencies of the nodes' total degrees  $\{d_i\}$  (i.e.,  $d_i = d_i^{(G)} + d_i^{(NG)}$ ), where the mean is  $\langle d_i \rangle \approx 2.59$ . (c) Histogram of the frequencies of the edge lengths  $\{\chi_{ij}\}$ , where  $\chi_{ij} = m(i, j)$  is the Euclidean distance between locations  $\mathbf{w}^{(i)}$  and  $\mathbf{w}^{(j)}$  for each edge  $(i, j) \in \mathcal{E}$  [see Eq. (16)]. See Supplementary Note 1 for further discussion.



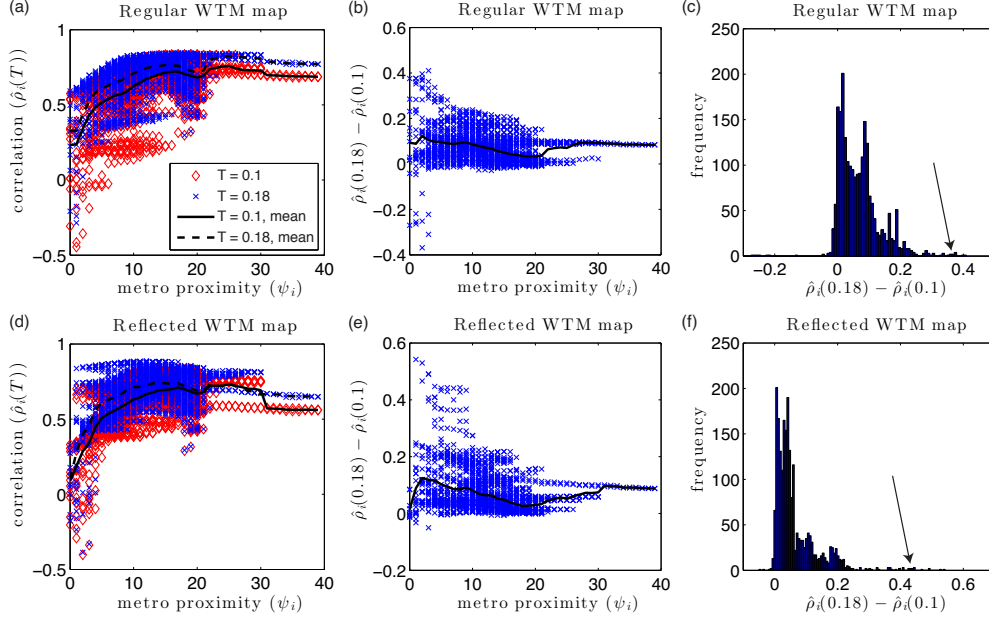
Supplementary Figure 2: Activation times  $\{x_j^{(i)}\}$  for nodes  $i \in \mathcal{V}$  for a WTM contagion on the London transit network, which we initiate with cluster seeding centered at a node  $j$  near the Bond Street Station. (a) For small thresholds, such as  $T = 0.02$ , nodes near metro stations have small activation times, so the contagion does not follow the geometric edges (i.e., the roads). (b) For moderate threshold values, such as  $T = 0.18$ , the activation times have a large positive correlation with the Euclidean distances between the intrinsic node locations  $\{w^{(i)}\}$  (given by latitude and longitude). Therefore, the WFP and ANC phenomena of WTM contagions with this initialization depend significantly on the value of  $T$ . Although this is not “typical” of all WTM contagions on this network, such situations have a significant effect on the resulting WTM maps. See Supplementary Note 1 for further discussion.



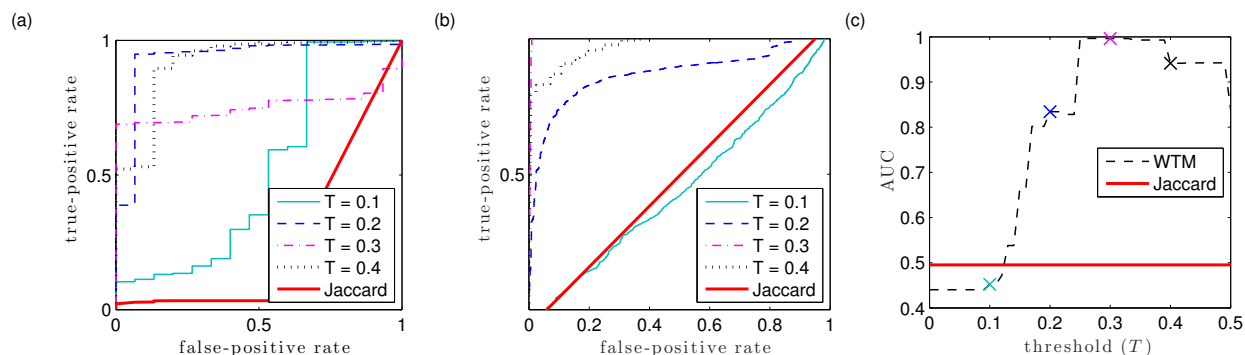
Supplementary Figure 3: Equilibrium contagion sizes and geometry of WTM maps on the London transit network. (a) Histogram of the frequency of sizes  $C_i^{(\text{target})}$  for the target node sets and sizes  $C_i^{(\text{source})}$  for the source node sets. Nodes tend to have either very large or very sizes of the target and source node sets, so we assign nodes into four classes: (1) large  $C_i^{(\text{target})}$  and large  $C_i^{(\text{source})}$ , (2) small  $C_i^{(\text{target})}$  and large  $C_i^{(\text{source})}$ , (3) large  $C_i^{(\text{target})}$  and small  $C_i^{(\text{source})}$ , and (4) small  $C_i^{(\text{target})}$  and small  $C_i^{(\text{source})}$ . In practice (as we discuss in Supplementary Note 1), we compare these values to  $N/2$  to assign nodes to classes. (b) Fraction of nodes in classes (1)–(4). All nodes shift from class (1) to class (4) as  $T$  increases; however, for the approximate range  $T \in (0.1, 0.25)$ , nodes are only in classes (1)–(3). (c) Pearson correlation coefficient  $\rho$  for the WTM map (solid curves), Isomap (horizontal dotted line), and a 2D Laplacian eigenmap (horizontal dashed line). For the WTM map, we show results for the regular (“reg”), reflected (“ref”), and symmetric (“sym”) versions of the WTM map. For each version, we handle the activation times of infinity in two ways: we either (1) set these activation times to be  $2N$  and consider the complete matrix of activation times (“full”) as we proceeded with our studies of synthetic networks; or (2) we neglect these values and examine only the remaining submatrix of activation times (“part”) after removing appropriate rows and columns. For the values of  $T$  for which nodes are exclusively in classes (1) and (2) [i.e., for  $T$  in the approximate range  $(0.1, 0.2)$ ], we find that  $\rho$  increases for the WTM maps when we neglect the activation times of infinity. For the WTM maps in which we set the activation times of infinity to  $2N$ , the values of  $\rho$  for  $T \gtrsim 0.1$  are considerably smaller than those for  $T \lesssim 0.1$ . This is especially prominent in the symmetric and reflected WTM maps. See Supplementary Note 1 for further discussion.



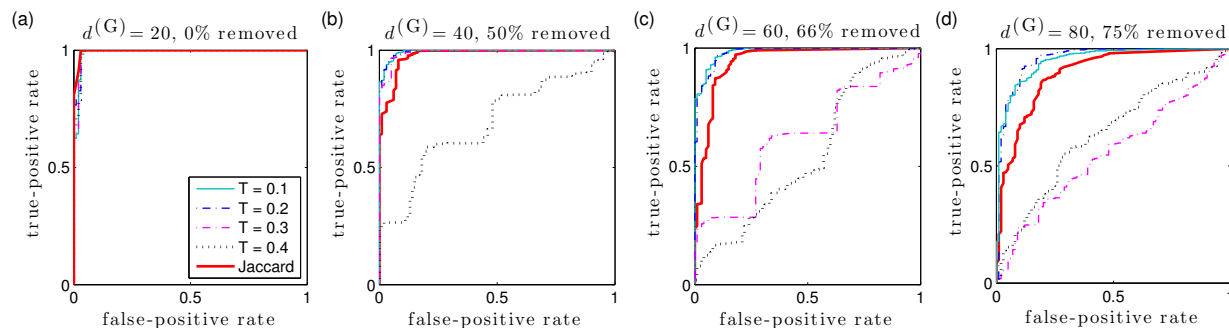
Supplementary Figure 4: (Top row) For a given value of the metro proximity  $\psi_i$  (i.e., the length of the shortest path from a node  $i$  to a metro station), we show the fraction of nodes that we assign to the four classes (1)–(4) based on the sizes of their target and source node sets (see the text). (a) For  $T = 0.16$ , almost all nodes are in class (1), though a few nodes are in class (2). The latter are located  $\psi_i = 2$  edges from a metro station. (b) For  $T = 0.18$ , most nodes are in class (1), but some nodes are in class (2). These are either 2–3 edges from a metro station, or they are “isolated” nodes that are distant from the other nodes (including, by definition, metro stations). (c) For  $T = 0.2$ , nodes are in classes (1)–(3). As before, nodes in class (2) are either 2–3 edges from a metro station or are isolated. The nodes in class (3) are relatively isolated. (Bottom row) Properties of the metro proximities  $\{\psi_i\}$ . (d) Frequency of nodes with a given metro proximity  $\psi_i$ . (e) Scatter plot of the nodes’ total degrees  $\{d_i\}$  versus their metro proximities  $\{\psi_i\}$ . (f) For each node  $i$ , we plot the mean length of the shortest paths from that node to the remaining nodes versus its metro proximity  $\psi_i$ . Note that isolated nodes, which are by definition distant from metro stations, are also distant from other nodes. See Supplementary Note 1 for further discussion.



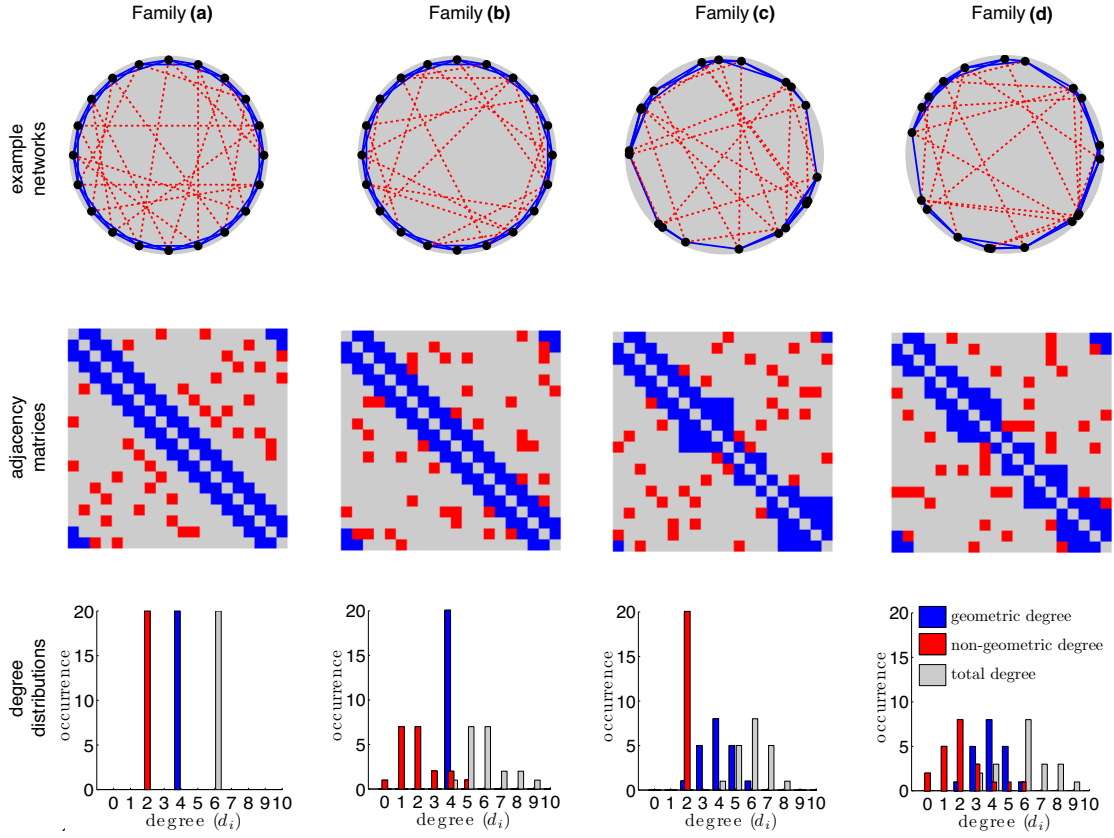
Supplementary Figure 5: Egocentric correlation coefficients  $\{\hat{\rho}_i(T)\}$  for WTM maps applied to the London transit network. We show results for (top panels) the map  $\mathcal{V}sto\{\mathbf{x}^{(i)}\}$  and (bottom panels) the map  $\mathcal{V}sto\{\mathbf{y}^{(i)}\}$ . For both maps, we neglect activation times that are infinite (see Supplementary Note 1). (Left column) In panels (a) and (d), we plot the egocentric correlations  $\{\hat{\rho}_i(T)\}$  versus the metro proximities  $\{\psi_i\}$  for thresholds  $T \in \{0.1, 0.18\}$ . We only show values for nodes that are not in class (2). The solid curve indicates the mean of  $\hat{\rho}_i(0.1)$  for a given value of  $\psi_i$ , and the dashed curve indicates the mean of  $\hat{\rho}_i(0.18)$  for a given value of  $\psi_i$ . Note that increasing  $T$  typically leads to an increase of  $\hat{\rho}_i(T)$  in both panels (a) and (d). (Center column) We plot  $[\hat{\rho}_i(0.18) - \hat{\rho}_i(0.1)]$  versus  $\psi_i$ . The solid curves indicate the mean values for a given  $\psi_i$ . Observe that the values of  $\hat{\rho}_i(T)$  are typically larger for  $T = 0.18$  than they are for  $T = 0.1$ . (Right column) We plot the observed frequencies of  $[\hat{\rho}_i(0.18) - \hat{\rho}_i(0.1)]$  for nodes  $j \in \mathcal{V}$ . Note that the frequencies are rather heterogeneous, and they appear to have a heavy tail (see the arrows): although  $[\hat{\rho}_i(0.18) - \hat{\rho}_i(0.1)]$  tends to be small and positive for most nodes  $j \in \mathcal{V}$ , there are some nodes for which  $[\hat{\rho}_i(0.18) - \hat{\rho}_i(0.1)]$  is rather large. See Supplementary Note 1 for further discussion.



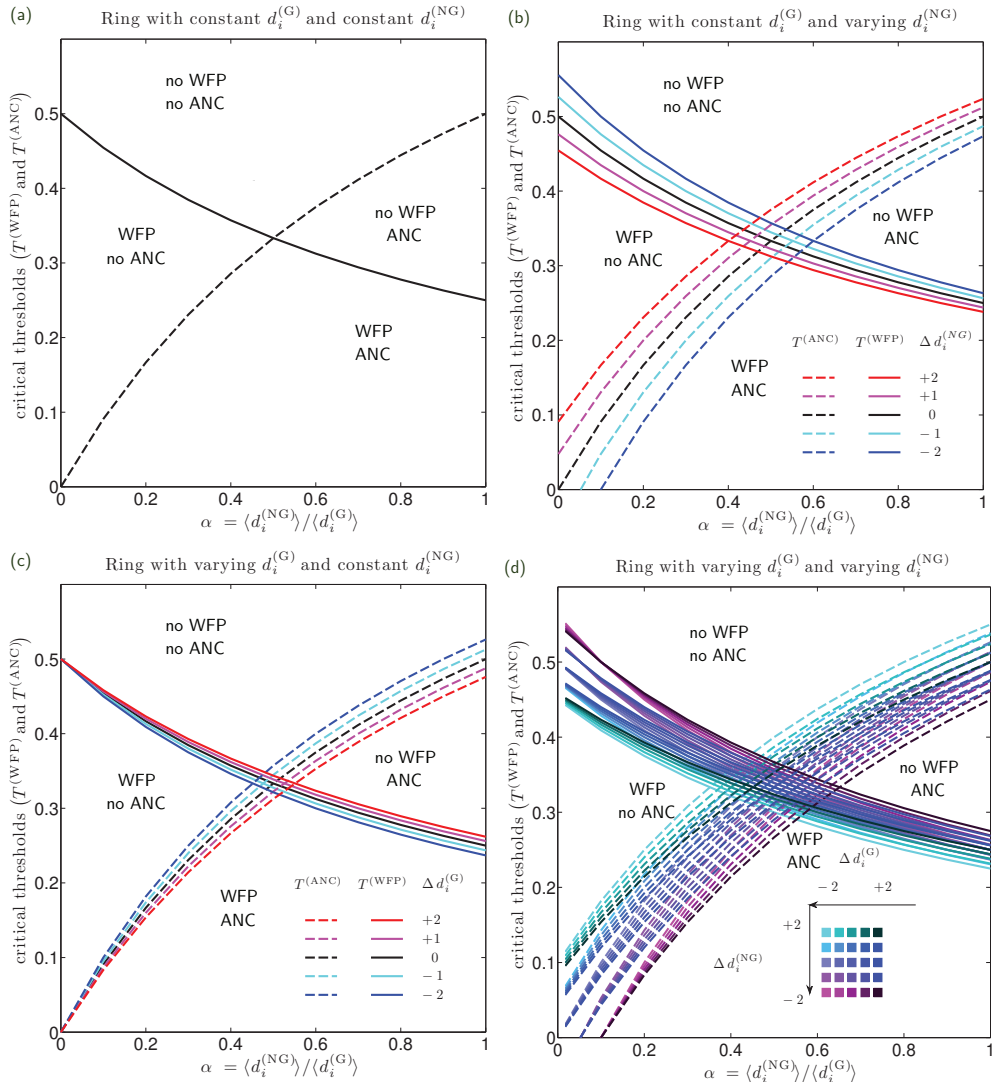
Supplementary Figure 6: Inference of noisy edges in the London transit network that we studied in Sec. I F of the main manuscript. We show results for WTM maps using various values of the threshold parameter  $T$ . (a) We show receiver operating characteristic (ROC) curves for the classification of edges in the London transit network as geometric or non-geometric based on WTM maps in addition to a “local” approach based on the Jaccard index [4]. (b) ROC curves for inferring the noisy edges in a noisy 2D square lattice. (c) We plot the the areas under the ROC curves (AUC) for various values of  $T$ . We use crosses to indicate the values of  $T$  that we used in panel (b). See Supplementary Note 2 for further discussion.



Supplementary Figure 7: ROC curves for the classification of edges as geometric or non-geometric based on WTM maps with various WTM thresholds  $T$ . The four panels (a)–(d) indicate results for four networks, which we generate so that their geometric edges have a tunable amount of stochasticity, which we implement by creating geometric edges and then removing some percentage of them uniformly at random. Note that an increased amount of stochasticity generally decreases the inference accuracy. See Supplementary Note 2 for further discussion.

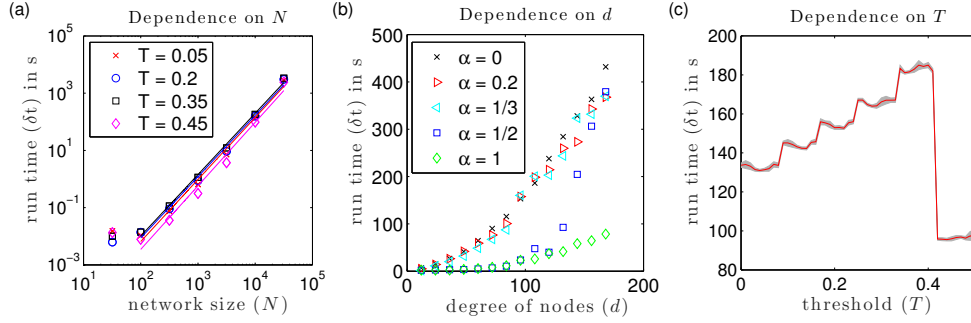


Supplementary Figure 8: Example networks with  $N = 20$  nodes, a mean geometric degree of  $\langle d_i^{(G)} \rangle = 4$ , and a mean non-geometric degree of  $\langle d_i^{(NG)} \rangle = 2$ , for four families of noisy geometric networks on a ring manifold: family **(a)**, the noisy ring lattice (which we also discuss in the main manuscript), for which nodes are evenly spaced and have constant geometric and non-geometric degrees; family **(b)**, for which the nodes are evenly spaced, have constant geometric degrees, and have heterogeneous non-geometric degrees; family **(c)**, for which we sample the node locations from the unit circle in  $\mathbb{R}^2$  using a stochastic process (see the text), and the nodes have heterogeneous geometric degree and constant non-geometric degrees; and family **(d)**, for which we randomly sample the node locations from the unit circle in  $\mathbb{R}^2$ , and the nodes have heterogeneous geometric and non-geometric degrees. (As we discuss in the text, we do the sampling uniformly at random.) The top row depicts example networks, where blue solid and red dashed lines indicate geometric and non-geometric edges, respectively. The center row depicts the corresponding adjacency matrices; blue pixels indicate geometric edges that align along the diagonal, whereas red pixels indicate non-geometric edges that arise randomly. The bottom row depicts the corresponding distributions for the geometric (red), non-geometric (blue), and total (grey) degrees. Note that the geometric degrees are identical for families **(a)** and **(b)**, with  $d_i^{(G)} = 4$ , whereas they are heterogeneous with mean  $\langle d_i^{(G)} \rangle = 4$  for families **(c)** and **(d)**. For families **(a)** and **(c)**, the non-geometric degrees are identical ( $d_i^{(NG)} = 2$ ), whereas they are heterogeneous with mean  $\langle d_i^{(NG)} \rangle = 2$  for families **(b)** and **(d)**. See Supplementary Note 3 for further discussion.

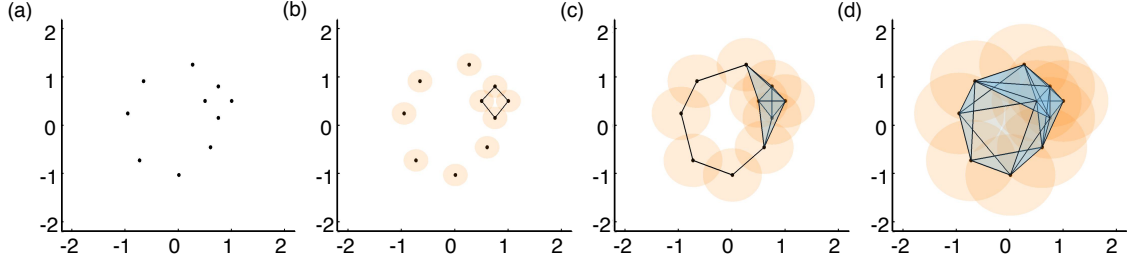


Supplementary Figure 9: Node-specific critical thresholds for the  $(T, \alpha)$  parameter plane, where  $\alpha$  denotes the ratio of the mean non-geometric degree  $\langle d_i^{(NG)} \rangle$  to the mean geometric degree  $\langle d_i^{(G)} \rangle$ . We plot Eqs. (3) and (4) for nodes with  $|\delta_i^{(G)}| \leq 2$  and  $|\delta_i^{(NG)}| \leq 2$ . We show these curves for the example degrees  $d_i^{(G)} = 20$  and  $d_i^{(NG)} \in [0, 20]$ . Panels (a)–(d), respectively, depict the appropriate choices for  $\delta_i^{(G)}$  and  $\delta_i^{(NG)}$  for the heterogeneities in network families (a)–(d). If the perturbations of the node degrees are both sufficiently small compared to the nodes’ degrees, then we still obtain four qualitatively different contagion regimes for all nodes. We again characterize these different regimes based on the presence versus absence of WFP and ANC. However, because of the heterogeneity of the nodes’ degrees, transitions between these regimes in the  $(T, \alpha)$  parameter plane occur at different values for different nodes. That is, the boundaries between the WTM contagion regimes have “thickened.” We note for any fixed  $|\delta^{(G)}|, |\delta^{(NG)}| > 0$  that the perturbed curves approach those that correspond to  $\delta^{(G)} = \delta^{(NG)} = 0$  as the mean geometric and non-geometric degrees increase. See Supplementary Note 3 for further discussion.

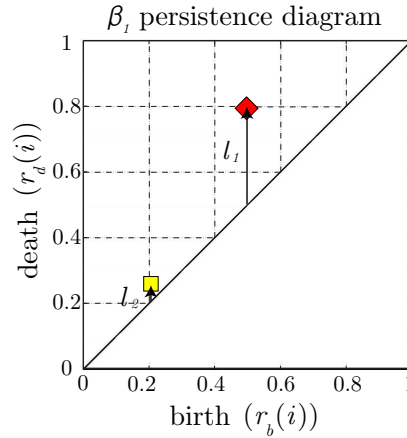




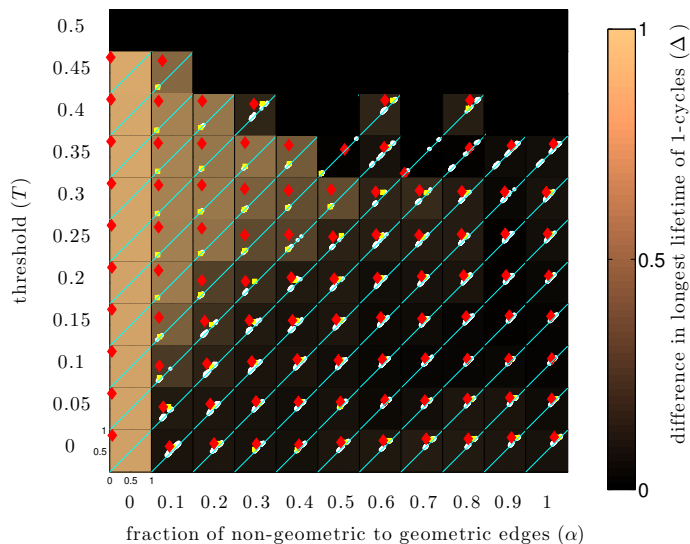
Supplementary Figure 10: Computational costs from experiments with Supplementary Algorithm 1. We use the run time  $\delta t$  (in seconds) to measure the cost of constructing WTM maps with threshold  $T$  for noisy ring lattices with  $N$  nodes. (a) We plot the observed values of  $\delta t$  (symbols) versus  $N$  for several choices of  $T$ . Note that  $\delta t$  varies only slightly with respect to  $T$ , whereas the dependence on  $N$  is much stronger. We show results for  $(d^{(G)}, d^{(NG)}) = (10, 2)$ ; that is,  $\alpha = 0.2$  and  $d = 12$ . The solid lines indicate the inferred scaling behaviors  $\mathcal{O}(N^\zeta)$ ; as we illustrate in Supplementary Table 1, the scalings are approximately quadratic  $\mathcal{O}(N^2)$ . (b) We plot observed values of  $\delta t$  for a noisy ring lattice with  $N = 2000$  nodes with various choices for the node degree  $d = d^{(G)} + d^{(NG)}$ . We show results for various  $d$  for several choices of the ratio  $\alpha = d^{(NG)}/d^{(NG)}$ . For all values of  $\alpha$ , we observe that  $\delta t$  scales approximately linearly with  $d$ . Networks with large values of  $\alpha$  promote transmission via ANC, which saturates the network in fewer time steps than that for smaller  $\alpha$ , which subsequently leads to considerably smaller run times (e.g., see the results for  $\alpha \geq 1/2$  versus  $\alpha < 1/2$ ). (c) The solid curve indicates  $\delta t$  versus the WTM threshold  $T$ ; the shaded region near the curve indicates the standard deviation over 10 realizations for a given threshold  $T$ . When a contagion saturates the network, so that all nodes eventually adopt the contagion (i.e.,  $T \leq T_0^{(\text{WFP})} \approx 0.4167$ ), observe that  $\delta t$  tends to increase with  $T$ . By contrast, when a contagion does not spread (i.e.,  $T \gtrsim 0.4167$ ), then  $\delta t$  is very small compared to the values when the contagion does spread. We also note that the abrupt jumps in  $\delta t$  are well-aligned with the critical thresholds given by Eqs. (7) and (13). The shaded region near the curve indicates the standard deviation (in units of time  $\delta t$ ) over 10 realizations for a given threshold  $T$ . See Supplementary Note 5 for further discussion.



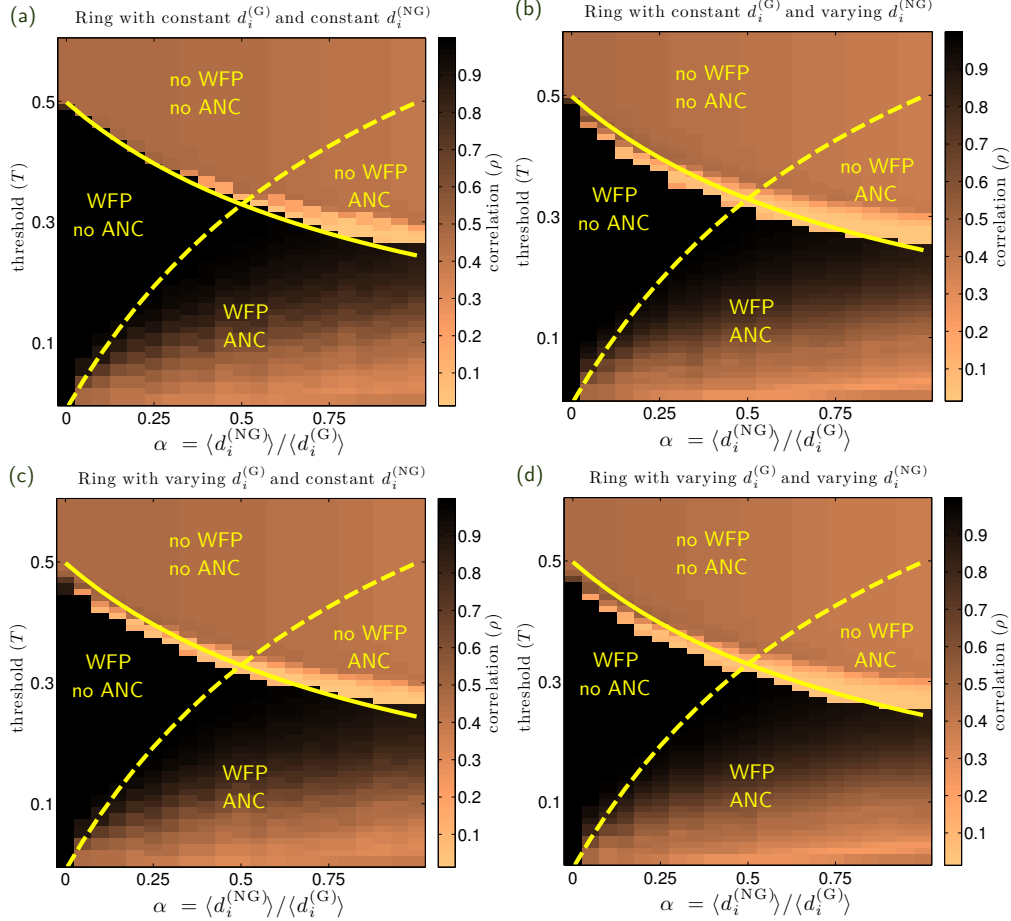
Supplementary Figure 11: We study the topology of a point cloud  $\mathcal{U}$  by examining the persistent homology that is induced by a Vietoris-Rips filtration. This entails examining simplicial complexes that are created by forming, for every set of points, a simplex (e.g., an edge, a triangle, a tetrahedron, etc.) whose diameter is at most  $r$ . Increasing  $r$  from 0 and considering how a simplicial complex evolves yields a filtration. In panel (a), we show a point cloud  $\mathcal{U} = \{\mathbf{u}^{(i)}\}$  that consists of a noisy sample of the unit circle. In this example, there are  $n = 10$  points in  $J = 2$  dimensions. In panels (b)–(d), we show  $\mathcal{U}^{(r)}$  for  $r \in \{0.22, 0.6, 0.85\}$ . One can approximate the homology of  $\mathcal{U}^{(r)}$  using a Vietoris-Rips complex that is given by the nodes, edges, and triangles that we show in the panels. The first 1-cycle in  $\mathcal{U}^{(r)}$  occurs at  $r = 0.22$ . It is a result of the noisy sampling, and it is filled in almost immediately as  $r$  increases. In panel (c), we show the dominant 1-cycle (i.e., the 1-cycle that corresponds to the ring and persists across many spatial scales). It is born at  $r = 0.5$  and persists until  $r \approx 0.81$ . Identifying a single persistent 1-cycle indicates that the point cloud lies on a ring manifold. See Supplementary Note 7 for further discussion.



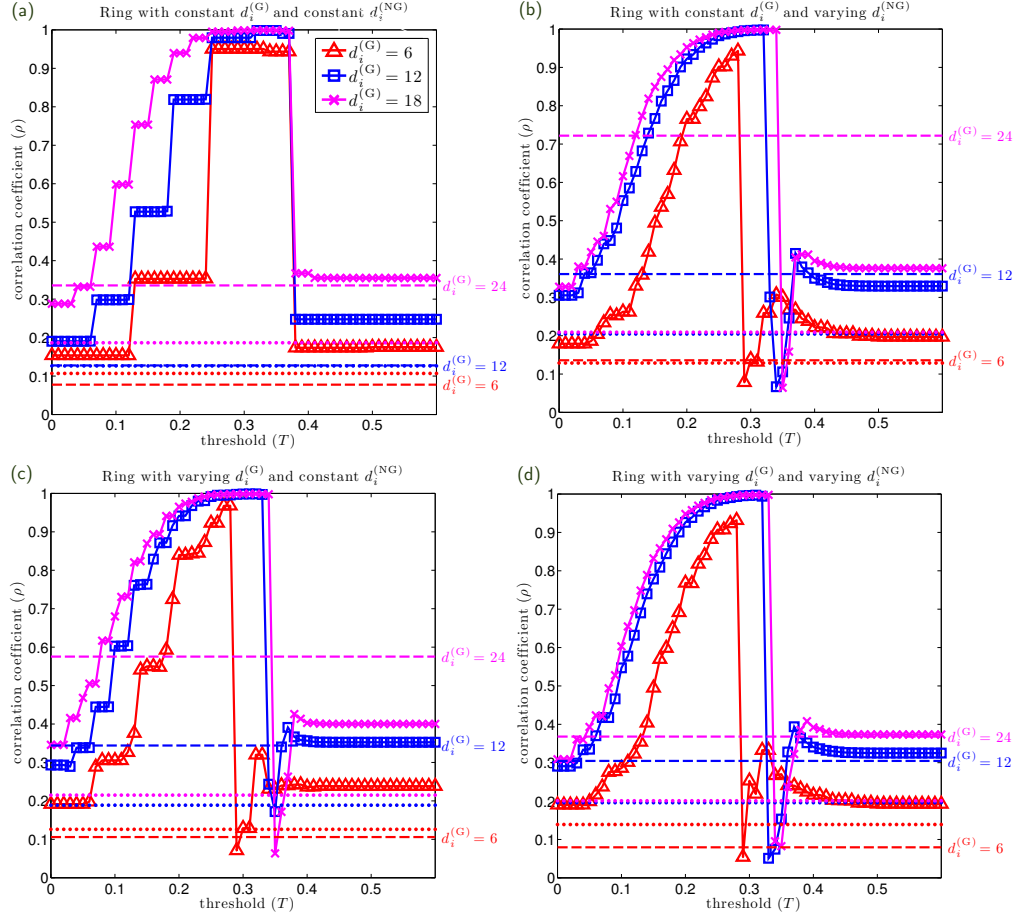
Supplementary Figure 12: A  $\beta_1$  persistence diagram that summarizes the 1D features (i.e., 1-cycles) that are revealed by the filtration  $\mathcal{U}^{(r)}$  in Supplementary Fig. 11. It contains two points, which correspond to the two observed 1-cycles. One point (the red diamond) indicates a 1-cycle that persists over a long range of spatial scales. Its lifetime  $l_1 = r_d(1) - r_b(1)$  is thus large. The second point (the yellow square) indicates another 1-cycle. Its small lifetime  $l_2 = r_d(2) - r_b(2)$  indicates that it dies a short time after it is born, so it does not persist over many spatial scales. The large difference  $\Delta = l_1 - l_2$  in the top two lifetimes indicates that the point cloud contains a single dominant 1-cycle and offers strong evidence that the point cloud lies on a ring manifold. See Supplementary Note 7 for further discussion.



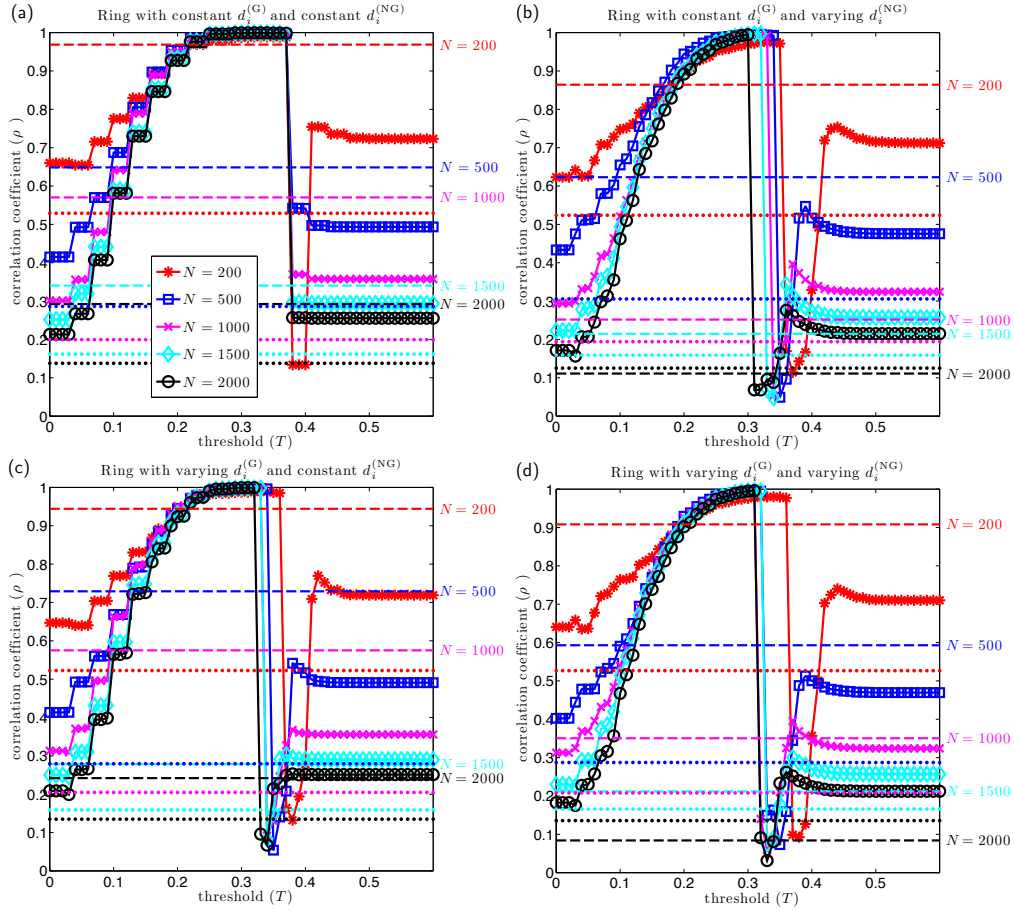
Supplementary Figure 13: We show a grid of  $\beta_1$  persistence diagrams for point clouds that result from the application of WTM maps with various values of the threshold  $T$  to noisy ring lattices for various values of the ratio  $\alpha = d^{(\text{NG})}/d^{(\text{G})}$  of non-geometric edges to geometric edges. We show results for  $T = 0, 0.05, \dots, 0.5$  for networks with  $N = 200$  nodes each with  $d^{(\text{G})} = 20$  geometric edges and  $d^{(\text{NG})} \in \{0, 2, \dots, 20\}$  non-geometric edges. For a given point cloud, we apply a Vietoris-Rips filtration to yield the  $\beta_1$  persistence diagram that summarizes the multiscale 1D features (i.e., 1-cycles or loops). In each persistence diagram, we use a red diamond to mark the most persistent 1-cycle, a yellow square to mark the second most persistent 1-cycle, and white circles to indicate the remaining 1-cycles that we find. Note that the lifetime  $l_i$  of a given point  $i$  corresponds to the height above the diagonal (the cyan lines). We shade the background color of each persistence diagram according to the difference  $\Delta = l_1 - l_2$  between the two largest lifetimes. Note that  $\Delta \in [0, 1]$  due to normalization. (See Sec. III F of the main manuscript and Supplementary Note 7.) The magnitude of  $\Delta$  provides strong evidence regarding whether or not a given point cloud arises from a 1D ring topology. In the main manuscript, we thus summarize our topological analysis with the parameter  $\Delta$ . [For example, see Fig. 6(c) of the main manuscript.] Note that we do not do any calculations (as indicated by the empty squares) for WTM maps in which any node has an activation time of infinity [i.e., when there is at least one pair  $(i, j)$  such that  $x_j^{(i)} = \infty$ ]. See Supplementary Note 7 for further discussion.



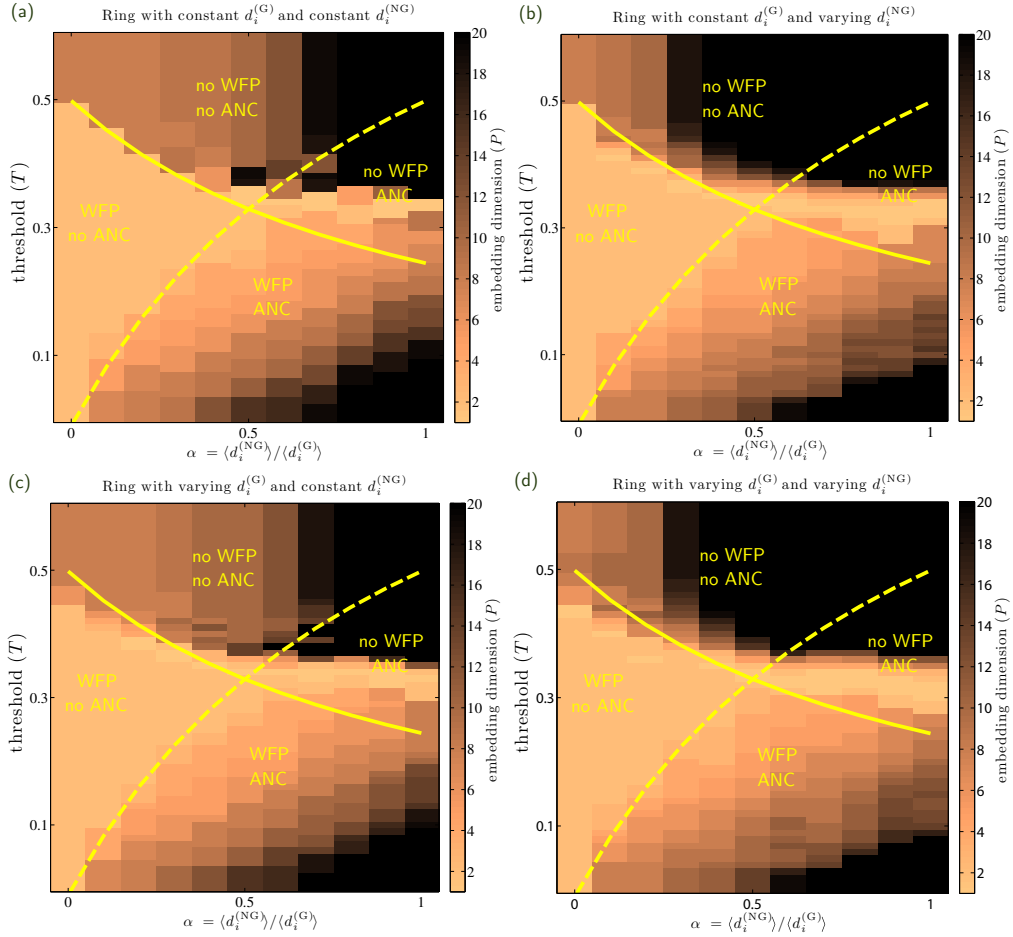
Supplementary Figure 14: We study the geometry of symmetric WTM maps by calculating a Pearson correlation coefficient  $\rho$  to compare node-to-node distances for the WTM map  $\{\mathbf{z}^{(i)}\} \in \mathbb{R}^N$  to those for the node locations  $\{\mathbf{w}^{(i)}\} \in \mathbb{R}^2$  on the ring manifold. (See Sec. III D of the main manuscript.) We show these values of  $\rho$  in the  $(T, \alpha)$  parameter plane, where  $\alpha$  is the ratio of the number  $N\langle d_i^{(NG)} \rangle / 2$  of non-geometric edges to the number  $N\langle d_i^{(G)} \rangle / 2$  of geometric edges. Panels (a)–(d), respectively, illustrate results for network families (a)–(d). For each panel, we construct a noisy ring network with  $N = 1000$  nodes and mean geometric degree  $\langle d_i^{(G)} \rangle = 40$ , and we vary the mean non-geometric degree  $\langle d_i^{(NG)} \rangle \in [0, 40]$  to study the parameter range  $\alpha \in [0, 1]$ . The solid and dashed curves, respectively, give the theoretical approximations from Eqs. (3) and (4) with  $\delta_i^{(G)} = \delta_i^{(NG)} = 0$ . See Supplementary Note 8 for further discussion.



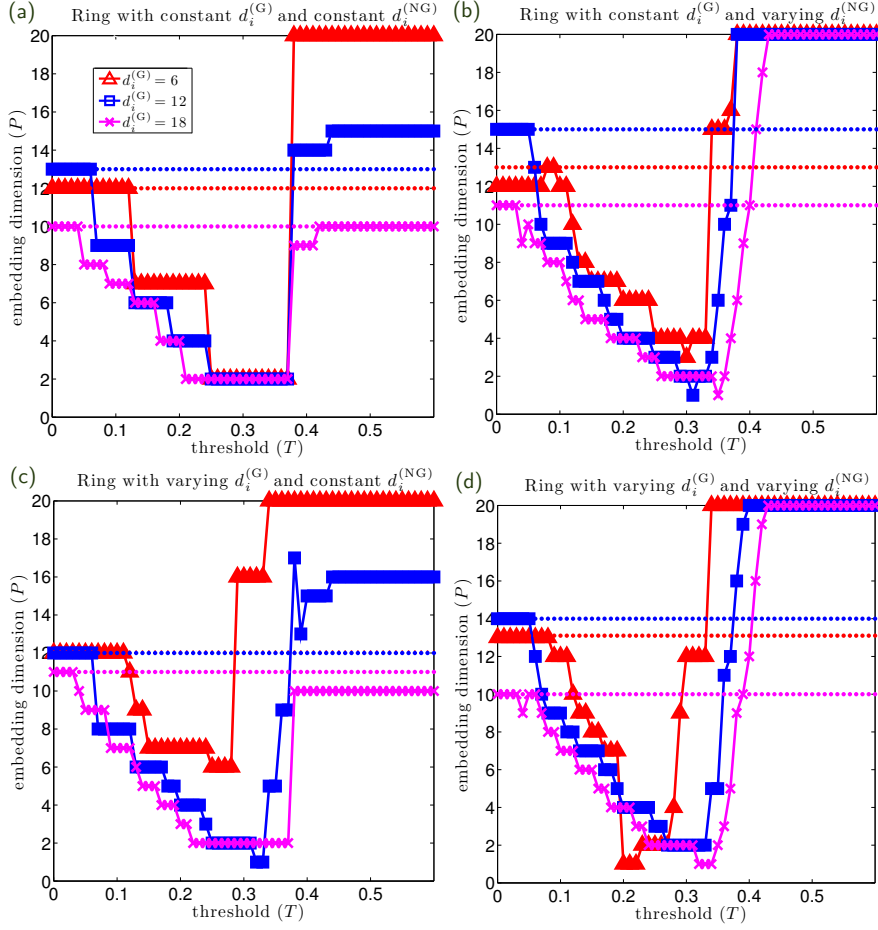
Supplementary Figure 15: We study the geometry of symmetric WTM maps by calculating a Pearson correlation coefficient  $\rho$  as a function of WTM threshold  $T$  to compare node-to-node distances for the WTM map  $\{\mathbf{z}^{(i)}\} \in \mathbb{R}^N$  to those for the node locations  $\{\mathbf{w}^{(i)}\} \in \mathbb{R}^2$  on a ring manifold. Panels (a)–(d), respectively, illustrate results for network families (a)–(d), and they amount to vertical cross sections of the corresponding contour plots in Supplementary Fig. 14 (i.e., for a constant value of  $\alpha$ ). In each panel, we study WTM maps on a noisy ring network with  $N = 1000$  nodes with  $\alpha = 1/3$  for several choices of mean node degrees:  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (6, 2)$  (red triangles),  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (12, 4)$  (blue squares), and  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (24, 8)$  (magenta  $\times$  symbols). We also plot  $\rho$  for a 2D Laplacian eigenmap [27] (dashed lines) and for the Isomap algorithm [23] (dotted lines). In all panels and for all mapping algorithms, increasing mean node degree tends to increase  $\rho$ , so the ability of the maps to translate the underlying ring manifold’s geometry to a point cloud improves with increasing mean node degree for these experiments. Additionally, note that the curves for the largest mean degree (magenta  $\times$  symbols) remain more consistent across the panels. See Supplementary Note 8 for further discussion.



Supplementary Figure 16: We study the geometry of symmetric WTM maps by calculating a Pearson correlation coefficient  $\rho$  as a function of WTM threshold  $T$  to compare node-to-node distances for the WTM map  $\{\mathbf{z}^{(i)}\} \in \mathbb{R}^N$  to those for the node locations  $\{\mathbf{w}^{(i)}\} \in \mathbb{R}^2$  on a ring manifold. Panels (a)–(d), respectively, illustrate results for network families (a)–(d), and they amount to vertical cross sections of the corresponding contour plots in Supplementary Fig. 14 (i.e., for a constant value of  $\alpha$ ), although we show results for several choices of network size  $N$ . In each panel, we study WTM maps on a noisy ring network with  $\langle d_i^{(G)} \rangle = 24$  and  $\langle d_i^{(NG)} \rangle = 8$  (i.e.,  $\alpha = 1/3$ ), and we show curves of  $\rho$  versus  $T$  for networks of sizes  $N = \{200, 500, 1000, 1500, 2000\}$ . We also plot  $\rho$  for a 2D Laplacian eigenmap [27] (dashed lines) and for the Isomap algorithm [23] (dotted lines). In all panels and for all mapping algorithms, increasing the network size  $N$  tends to decrease  $\rho$ , except for WTM maps that are characterized by WFP and little (or no) ANC. See Supplementary Note 8 for further discussion.

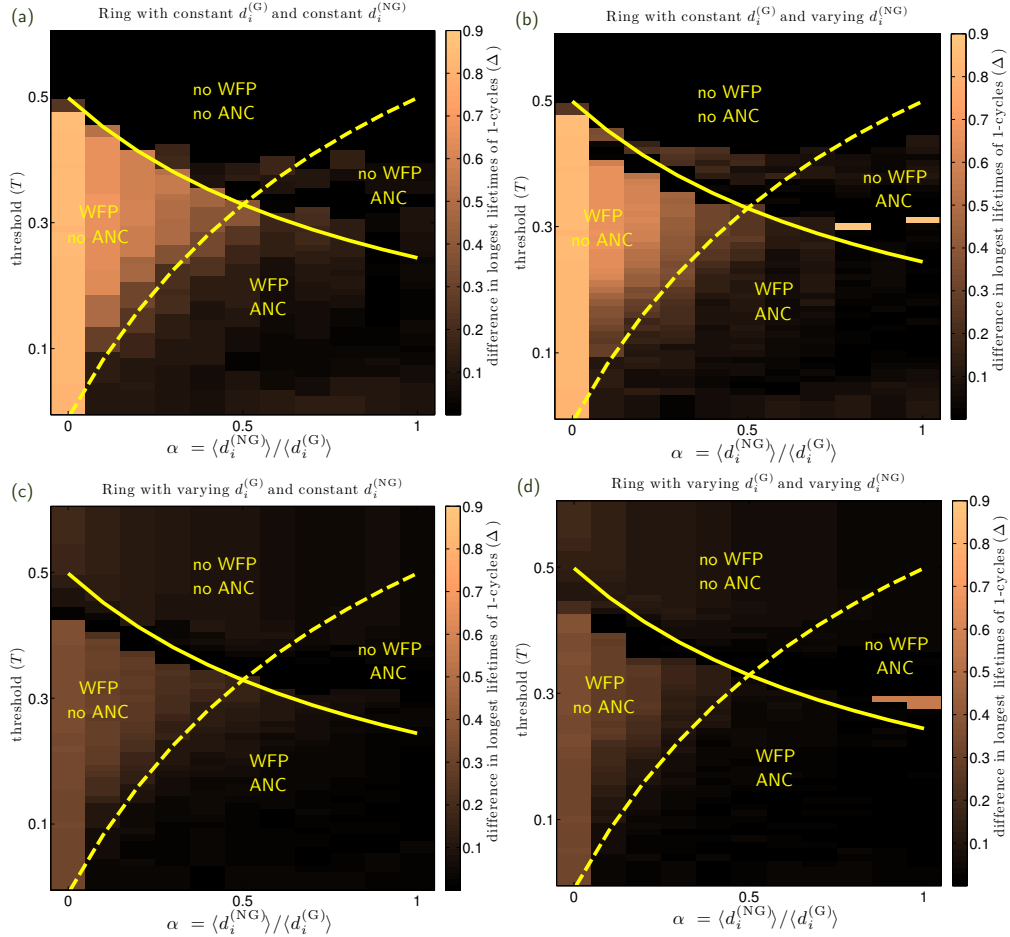


Supplementary Figure 17: We examine the dimensionality of point clouds that result from symmetric WTM maps that we apply to networks on a ring manifold by studying their embedding dimension  $P = \min\{p | R_p < 0.05\}$ , where  $R_p$  denotes the residual variance for the projection onto  $\mathbb{R}^p$ . (See Sec. III E of the main manuscript.) We plot  $P$  in the  $(T, \alpha)$  parameter plane for networks with  $N = 200$  nodes, mean geometric degree  $\langle d_i^{(G)} \rangle = 20$ , and various values of the non-geometric degree  $\langle d_i^{(NG)} \rangle$ . As before, panels (a)–(d), respectively, illustrate the results for network families (a)–(d). We use solid and dashed curves, respectively, to indicate the theoretical critical threshold values given by Eqs. (3) and (4) with  $\delta_i^{(G)} = \delta_i^{(NG)} = 0$ . In all panels, we see that WTM maps for the contagion regime that we predict to be characterized by WFP but no ANC yield point clouds with an embedding dimension of  $P \approx 2$ , which agrees with the fact that a ring manifold is embedded in  $\mathbb{R}^2$ . This low-dimensional structure persists into the regime that we predict has both WFP and ANC, although the embedding dimension  $P$  increases as one moves away from the regime that exhibits WFP and no ANC. See Supplementary Note 8 for further discussion.

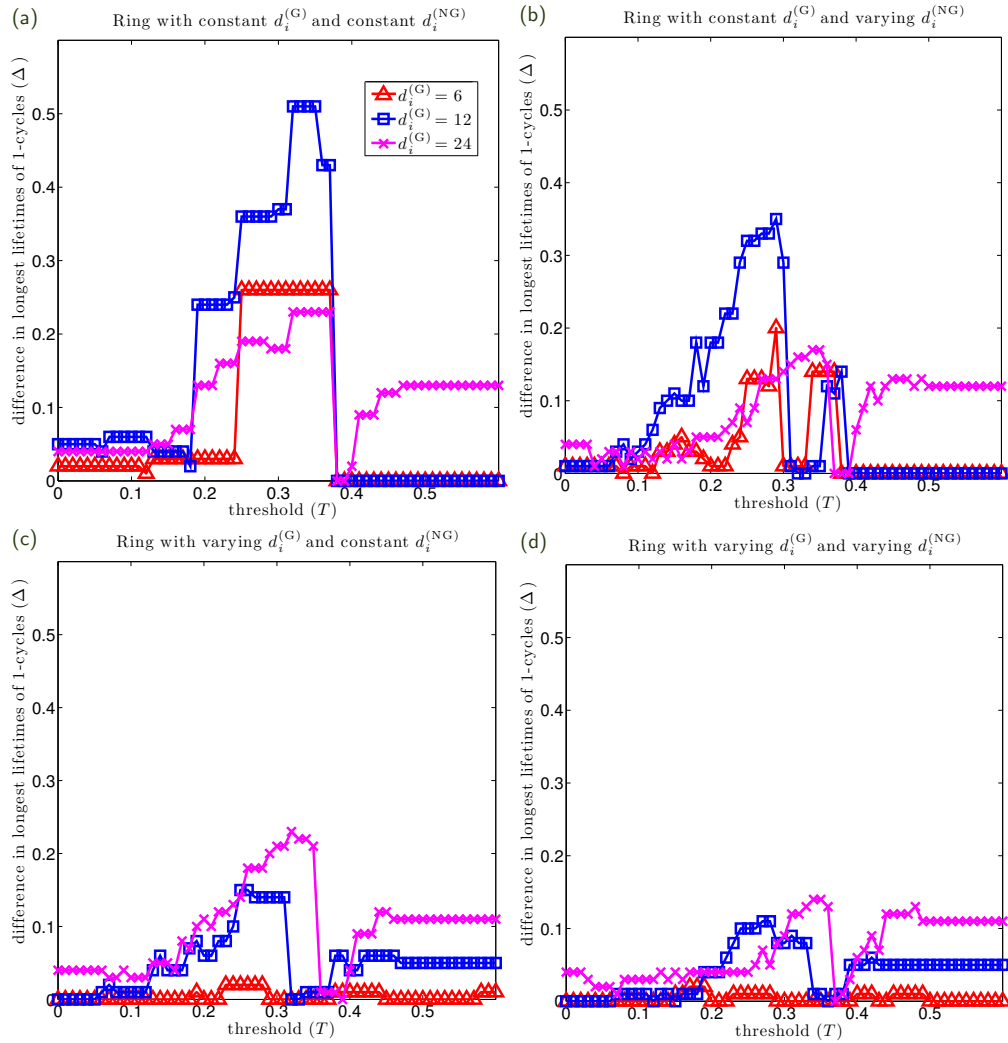


Supplementary Figure 18: We study the dimensionality of point clouds that result from symmetric WTM maps by showing their embedding dimension  $P$  as a function of  $T$  for networks with  $N = 200$  and  $\alpha = 1/3$ . One can construe these curves of  $P$  versus  $T$  as vertical cross sections for the contour plots in Supplementary Fig. 17; we show results for several choices of mean node degrees:  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (6, 2)$  (red triangles),  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (12, 4)$  (blue squares), and  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (18, 6)$  (magenta  $\times$  symbols). As before, panels (a)–(d) correspond to network families (a)–(d). In all panels, we identify the correct dimension (i.e.,  $P = 2$ ) for the regime that we expect WTM contagion to exhibit WFP without ANC [i.e.,  $T \in (1/4, 3/8)$ ] if the mean node degrees are sufficiently large (see magenta  $\times$  symbols). The curves for the other mean degrees also consistently depict small values of  $P$  for a similar range of threshold  $T$ . We also plot  $P$  versus  $T$  for the point clouds that we obtain by mapping the nodes based on shortest paths, as in Isomap [23] (horizontal dotted lines). For these experiments,  $P \geq 10$  from Isomap in all panels, although it appears to decrease systematically with increasing mean node degrees. See Supplementary Note 8 for further discussion.

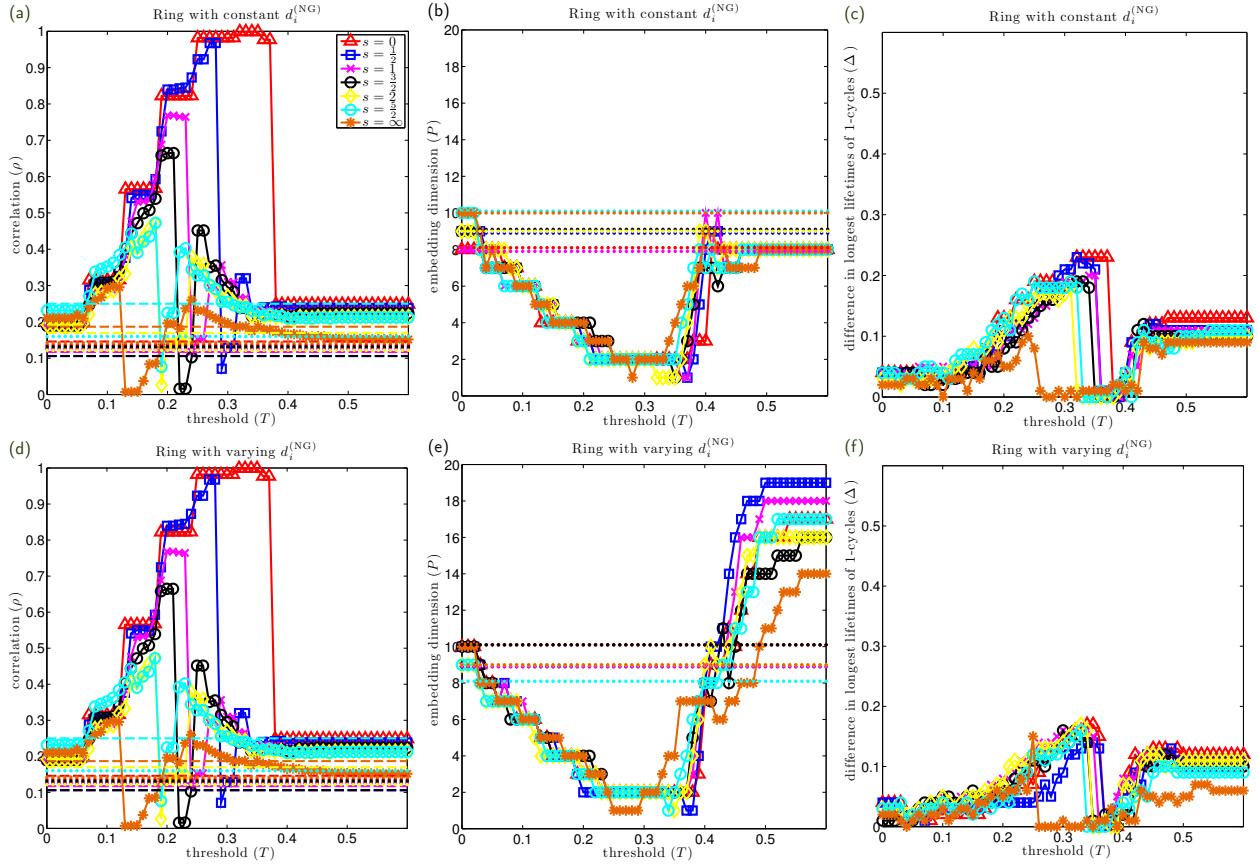




Supplementary Figure 19: We study the topology of symmetric WTM maps by calculating the difference  $\Delta = l_1 - l_2$  between the largest lifetimes for 1D features (i.e., for 1-cycles); see Sec. III F of the main manuscript. We normalize the lifetime difference so that  $\Delta \in [0, 1]$ . Panels (a)–(d), respectively, show results for network families (a)–(d). Each network has  $N = 200$  nodes, a mean geometric degree of  $\langle d_i^{(G)} \rangle = 20$ , and various mean non-geometric degrees  $\langle d_i^{(NG)} \rangle \in [0, 20]$ . The solid and dashed curves, respectively, give the approximate bifurcation curves from Eqs. (3) and (4) with  $\delta_i^{(G)} = \delta_i^{(NG)} = 0$ . Panel (a) is similar to Fig. 6(c) of the main manuscript. In all panels, we observe evidence that the point clouds that result from WTM maps that correspond to contagions with WFP but no ANC lie on a ring manifold. However, this evidence becomes weaker (as indicated by smaller values of  $\Delta$ ) as the networks become more heterogeneous [e.g., compare panel (d) to panel (a)]. See Supplementary Note 8 for further discussion.



Supplementary Figure 20: We study the topology of symmetric WTM maps by plotting  $\Delta$  as a function of  $T$  for  $N = 200$  and  $\alpha = 1/3$ . As before, panels (a)–(d) correspond, respectively, to network families (a)–(d). One can construe the curves of  $\Delta$  versus  $T$  as vertical cross sections of the contour plots in Supplementary Fig. 19; we consider several different choices of mean node degrees:  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (6, 2)$  (red triangles),  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (12, 4)$  (blue squares), and  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (24, 8)$  (magenta  $\times$  symbols). Note that introducing heterogeneity tends to decrease the ability to identify the ring topology in the point cloud with  $\Delta$ . For example, note that the values of  $\Delta$  in panels (b) and (c) are smaller than those in panel (a), and the values of  $\Delta$  in panel (d) are smaller than those in panels (b) and (c). Additionally, in panel (c) and panel (d), we see that when the mean degrees are too small (e.g., see red triangles), then  $\Delta \approx 0$  for all thresholds  $T$ . Thus, we do not find evidence of the ring topology for these point clouds. See Supplementary Note 8 for further discussion.



Supplementary Figure 21: We study the (left column) geometry, (center column) dimensionality, and (right column) topology for symmetric WTM maps applied to noisy geometric networks with node locations  $\{\mathbf{w}^{(i)}\}$  that we sample randomly from a ring manifold, where the parameter  $s$  controls the amount of heterogeneity that we introduce into the spacing of the nodes along the ring. (Recall that we used  $s = 1/2$  to define network families (c) and (d).) We show results for various values of the threshold  $T$  for networks with  $N = 200$  nodes, mean geometric degree  $\langle d_i^{(G)} \rangle = 24$ , and mean non-geometric degree  $\langle d_i^{(NG)} \rangle = 8$ . We consider both non-geometric degrees that are (top row) constant across the nodes, as in network family (c) and (bottom row) heterogeneous, as in family (d). Note that increasing heterogeneity in node spacing on the manifold, which in turn increases the heterogeneity in the geometric degrees  $\{d_i^{(G)}\}$  (although their mean is constant), tends to lead to a decrease in the ability of the symmetric WTM maps to recover the properties of the underlying manifold in the resulting point cloud. One sees this mostly clearly when examining the geometry and topology, as there are significant drops in  $\rho$  and  $\Delta$  as  $s$  increases. The dotted lines in panels (a), (b), (d), and (e) indicate the values that we observe for the mapping of nodes based on shortest paths, as in the Isomap algorithm. The dashed lines in panels (a) and (d) indicate values for a 2D Laplacian eigenmap. See Supplementary Note 8 for further discussion.

Inferred run time scaling:  $\delta t = 10^\Gamma N^\zeta$

threshold $T$	exponent $\zeta$	prefactor $\Gamma$
0.05	2.1675	-6.5207
0.2	2.1407	-6.3416
0.35	2.1440	-6.2792
0.45	2.2195	-6.8987

Supplementary Table 1: For various choices of threshold  $T$ , we infer the scaling behavior that relates the computational cost (i.e., the run time  $\delta t$ ) to the network size  $N$ . For our inference procedure, we assume a power-law relationship  $\delta t = 10^\Gamma N^\zeta$ , and we fit the constants  $\Gamma$  and  $\zeta$  using a least-squares fit. In this fit, the horizontal coordinates are  $\log(N)$ , and the vertical coordinates are  $\log(\delta t)$  (We neglect the results for  $N = 32$  in our fitting procedure.) Note that the exponents are approximately quadratic:  $\zeta \approx 2$ . See Supplementary Note 5 for further discussion.

## Supplementary Note 1 Complex Contagions on a London Transit Network

The primary goal of our work has been to develop the notion of a WTM map and to demonstrate the utility of using such maps for examining WTM contagions on noisy geometric networks. Specifically, we conducted a detailed examination that contrasts wavefront propagation (WFP) along geometric edges versus the appearance of new contagion clusters (ANC) due to the presence of non-geometric, “noisy” edges. We have focused on synthetic networks—and, in particular, on noisy geometric networks on a ring manifold—and we conducted a bifurcation analysis to guide our study. However, one can use WTM maps on far more general types of networks such as noisy geometric networks that are constructed from empirical data. (More generally, one can also use contagion dynamics that one constructs from other types of spreading processes.) This allows two important applications to real systems: (1) one can study the extent to which a contagion on a network exhibits spatial phenomena such as WFP versus non-spatial phenomena such as ANC; and (2) one can infer (potentially) unknown low-dimensional structure in a network. In this section, we highlight these ideas for an empirical network that describes transit infrastructure in a part of London.

### Description of the London Transit Network

As we illustrate in Supplementary Fig. 1(a), we study WTM contagions on a London transit network that includes both roads (which we interpret as short-range, geometric edges) and metro lines (which we interpret as long-range, non-geometric edges). The nodes  $\mathcal{V} = \{1, \dots, N\}$  (where  $N = 2217$ ) in the network correspond to intersections, and we obtain the edges from Refs. [1] (road data) and [2] (metro data). We construct the merged network (which we have posted, as we discussed in Sec. III A of the main text), by utilizing the latitudinal and the longitudinal coordinates to place the locations of metro stations at the nearest intersection of roads. Thus, the nodes  $\mathcal{V}$  consist of two sets: (1) nodes  $\mathcal{P} \subset \mathcal{V}$  that correspond to both metro stations and intersections and thus have both geometric and non-geometric edges; and (2) nodes  $\mathcal{V} \setminus \mathcal{P}$  that correspond to intersections and have only geometric edges (i.e., roads). Additionally, because the network of metro lines in Ref. [2] covers a much larger spatial area than the road network in Ref. [1], we include only metro stations in the convex hull of the road network. (There are  $|\mathcal{P}| = 11$  such stations.) In Supplementary Fig. 1(b), we show histograms of the frequencies of the nodes’ total degrees  $\{d_i\}$ , where  $d_i = d_i^{(G)} + d_i^{(NG)}$  and the mean is  $\langle d_i \rangle \approx 2.59$ . In Supplementary Fig. 1(c), we show histograms of the frequencies of the edge lengths  $\{\chi_{ij}\}$ , where  $\chi_{ij} = m(i, j)$  is the Euclidean distance between locations  $\mathbf{w}^{(i)}$  and  $\mathbf{w}^{(j)}$  [see Eq. (16)] for each edge  $(i, j) \in \mathcal{E}$ . In practice, we give node  $i$  an intrinsic location of  $\mathbf{w}^{(i)} = [w_1^{(i)}, w_2^{(i)}]^T$ , where  $w_1^{(i)}$  and  $w_2^{(i)}$  denote, respectively, the intersection’s latitudinal and longitudinal coordinate. We normalize each set of coordinates to have unit variance [see Supplementary Fig. 1(a)]. In general, such a projection from a patch on the surface of a sphere (e.g., the Earth’s surface) to a 2D plane might not be justified. However, the effect of this projection to a plane is negligible in this case due to the very small size of the patch.

Before analyzing WTM contagions and WTM maps for the London transit network, let’s consider the following experiment. In Supplementary Fig. 2, we illustrate that the extent to which a WTM contagion adheres to the network’s underlying manifold—the Earth’s surface—can be very sensitive to a variety of factors, including the contagion seed and the WTM threshold  $T$ . We plot the London transit network and color each node  $i \in \mathcal{V}$  according to its activation time  $x_j^{(i)}$  for a single contagion that we initialize with cluster seeding centered at a node  $j$ , which we take to be near the Bond Street Station. In panels (a) and (b), we show  $\{x_j^{(i)}\}$  for nodes  $i \in \mathcal{V}$  with thresholds of  $T = 0.02$  and  $T = 0.18$ , respectively. Note for  $T = 0.02$  that the contagion spreads via both roads and metro lines, so the contagion includes ANC. By contrast,

for  $T = 0.18$ , the contagion does not spread across the metro lines; rather, it spreads via WFP along the roads. As we shall see, this extreme sensitivity to the threshold  $T$  for the behavior of WTM contagions is not typical for all contagion seeds. Nevertheless, we find that such rare cases can have a large impact on the network’s WTM maps.

## Numerical Results for the Geometry of WTM Maps

In this section, we study the geometry of WTM contagions on the London transit network that we studied in Sec. I F of the main text by examining the geometry of WTM maps. As before, we study geometry through the Pearson correlation coefficient  $\rho$  given by Eq. (18). We do not study the dimensionality and topology because of the large computational time that it would entail.

To guide our investigation, we first study the equilibrium sizes of contagions (i.e., the number of infected nodes after the contagion stops spreading [3]). Our motivation is as follows. Recall that for WTM maps to be well-defined, all activation times  $\{x_j^{(i)}\}$  must be finite. In our numerical experiments for synthetic networks, we therefore focused on this situation (e.g., see the main text and Supplementary Note 8), and we chose to handle activation times that were infinite by setting them to be  $2N$ . Even with the restriction to finite activation times, we found a rich set of diverse qualitative dynamics. However, for the London transit network, the most interesting WTM maps occur for threshold values  $T$  that involve activation times of infinity. For this example, we must account for activation times of infinity more carefully to be able to study WTM contagions with WTM maps in such situations.

We thus begin by studying the equilibrium sizes of WTM contagions that we initialize with cluster seeding centered at each node  $i \in \mathcal{V}$ . Specifically, for a given threshold  $T$ , we study the size  $C_i^{(\text{target})}$  of each node  $i$ ’s “target node set” (which we define as the set of nodes  $\{j\}$  such that  $x_i^{(j)}$  is finite) and the size  $C_i^{(\text{source})}$  of its “source node set” (which we define as the set of nodes  $\{j\}$  such that  $x_j^{(i)}$  is finite). In other words, node  $j$  is in the target node set for node  $i$  if a contagion that is initialized at node  $i$  eventually spreads to node  $j$ , and node  $j$  is in the source node set for node  $i$  if a contagion that is handle at node  $j$  eventually spreads to node  $i$ .

In Supplementary Fig. 3(a), we show histograms of the frequencies of (top panel)  $C_i^{(\text{target})}$  and (bottom panel)  $C_i^{(\text{source})}$  for the network nodes for WTM contagions with threshold values of  $T \in \{0, 0.1, 0.2, 0.3\}$  that we initialize with cluster seeding. As expected, the WTM contagions infect almost all (or all) of the nodes when  $T$  is small, whereas they spread to just a small number of nodes (or even 0 nodes) when  $T$  is sufficiently large. For example, observe for most nodes that  $C_i^{(\text{target})}$  and  $C_i^{(\text{source})}$  are approximately  $N = 2217$  for  $T \leq 0.2$ , whereas they are very small for most nodes for  $T = 0.3$ . Additionally, the target and source node sets seem to exhibit dichotomous behavior in our experiments: they are often either very large (i.e., equal to or only a bit smaller than  $N$ ) or very small (i.e., approximately 1). We observe this feature both for the values of  $T$  that we depict as well as for other values of  $T$ . (We examined  $T \in \{0.01, 0.02, \dots, 0.5\}$ .) Motivated by this observation, we assign the nodes to four classes for a given  $T$ : (1) nodes  $i$  with large  $C_i^{(\text{target})}$  and large  $C_i^{(\text{source})}$  that can initiate large contagions and also adopt most contagions; (2) nodes  $i$  with small  $C_i^{(\text{target})}$  and large  $C_i^{(\text{source})}$  that do not initialize large contagions but adopt almost all contagions; (3) nodes  $i$  with large  $C_i^{(\text{target})}$  and small  $C_i^{(\text{source})}$  that initialize large contagions but almost never adopt contagions; and (4) nodes  $i$  with small  $C_i^{(\text{target})}$  and small  $C_i^{(\text{source})}$  that neither initialize large contagions nor adopt many contagions. In this classification, we arbitrarily take  $N/2$  to be the boundary between “large” and “small” for both types of division.

In Supplementary Fig. 3(b), we examine the fraction of nodes in each class as a function of  $T$ . For sufficiently small  $T$  (e.g.,  $T < 0.1$ ), almost all network nodes are in class (1) and almost all WTM contagions saturate the entire network. However, for large  $T$  (e.g.,  $T > 0.35$ ), all nodes are in class (4) because no contagions spread if the threshold is sufficiently large. The transitions between the different classes are interesting. Specifically, observe for the approximate range  $T \in (0.1, 0.2)$  that a small fraction of nodes moves from class (1) to class (2). Moreover, for the approximate range  $T \in (0.2, 0.25)$ , class (2) and class (3) each contain only a small fraction of the nodes. Class (4) remains empty until  $T \geq 0.25$ , and it then grows as we increase  $T$  until all nodes are in class (4) for  $T \gtrsim 0.35$ .

In Supplementary Fig. 3(c), we plot the Pearson correlation coefficient  $\rho$  from Eq. (18) to compare the geometry of the nodes' original locations  $\{\mathbf{w}^{(i)}\}$  to point clouds that result from WTM maps. We show results for the regular, reflected, and symmetric versions of the WTM map. (See Sec. I C of the main text.) We consider two different methods for handling the activation times of infinity [which necessarily arise whenever nodes are in classes (2)–(4)]. We either set the activation times to  $2N$  and investigate the complete matrix of activation times, or we consider only finite activation times by using only the associated submatrix of activation times (after removing appropriate rows and columns that contain activation times of infinity). To illustrate our analysis, consider the latter case for the map  $\mathcal{V}sto\{\mathbf{x}^{(i)}\}$ . We project each point  $\mathbf{x}^{(i)} \in \mathbb{R}^N$  onto  $\mathbb{R}^J$  with  $J \leq N$  by ignoring the dimensions that correspond to WTM contagions that are initialized with cluster seeding at nodes in class (2). This corresponds to considering the point cloud  $\{\hat{\mathbf{x}}^{(i)}\}$ , where  $\hat{\mathbf{x}}^{(i)} = \Omega \mathbf{x}^{(i)}$  and the  $J \times N$  projection matrix  $\Omega$  has entries  $\Omega_{jk_j} = 1$ , where  $j \in \{1, \dots, J\}$ , the set  $\{k_1, k_2, \dots, k_J\}$  indicates the nodes that are not in class (2), and all other entries  $\Omega_{jk}$  are equal to 0. For the reflected WTM map, we consider the map  $\{i\}sto\{\mathbf{y}^{(i)}\}$  only for nodes  $i$  that are not in class (2). Finally, for the symmetric WTM map, we consider the map  $\{i\}sto\{\hat{\mathbf{z}}^{(i)}\}$ , where  $\hat{\mathbf{z}}^{(i)} = \Omega \mathbf{z}^{(i)}$ , and we only map nodes  $i \in \mathcal{V}$  that are not in class (2).

Returning our attention to Supplementary Fig. 3(c), note for the reflected and symmetric WTM maps that we calculate the Pearson correlation coefficients  $\rho$  only for the mapped points. As expected,  $\rho$  for the WTM maps depends significantly on  $T$ , and one can observe that shifts in  $\rho$  are well-aligned with changes in  $C_i^{(\text{target})}$  and  $C_i^{(\text{source})}$ . The approximate range of thresholds  $T \in (0.1, 0.2)$  is particularly interesting, as we observe that values of  $\rho$  for WTM maps increase when we neglect the activation times that are infinite. These larger  $\rho$  values, in turn, indicate an improved agreement between the nodes' original locations and the geometry of the point clouds that result from the WTM maps. By contrast, for WTM maps in which we handle activation times of infinity by setting them to  $2N$ , we find that the values of  $\rho$  are smaller for  $T \gtrsim 0.1$  than they are for  $T \lesssim 0.1$ . That is, when we handle the activation times of infinity in this way, we find that the WTM map becomes significantly distorted away from the known spatial embedding on Earth's surface.

We now attempt to gain some insight into which nodes we assign to classes (1)–(4). In Supplementary Fig. 4, we investigate the importance of the nodes' *metro proximities*  $\{\psi_i\}$ , where  $\psi_i$  denotes the length of a shortest path on the London transit network from node  $i$  to a metro station (i.e.,  $\psi_i = 0$  for nodes that are metro stations,  $\psi_i = 1$  for their neighbors, and so on). We consider nodes that are at least 20 edges from any metro station to be "isolated." In the top row, for a given value of the metro proximity  $\psi_i$ , we plot the fraction of nodes at that proximity in each of the four classes. Panels (a), (b), and (c), respectively, give results for threshold values of  $T = 0.16$ ,  $T = 0.18$ , and  $T = 0.2$ , which are characteristic of the range of  $T$  in which we observe large values of  $\rho$  for the WTM maps [see Supplementary Fig. 3(c)]. Note for  $T = 0.16$  that almost all nodes are in class (1), but several are in class (2). Interestingly, all nodes in class (2) are located  $\psi_i = 2$  edges from metro stations. It follows that a WTM contagion tends not to spread very far when we initialize it with cluster seeding centered at such nodes. For  $T = 0.18$ , we again find that some nodes are in class (2), whereas the majority of nodes are in class (1). However, the nodes in class (2) are either 2–3

edges from a metro station or they are isolated nodes, which are distant from all other nodes (including, by definition, metro stations). For  $T = 0.2$ , we find that nodes are in classes (1)–(3). As before, nodes in class (2) are either 2–3 edges from a metro station or are isolated. The nodes in class (3)—which are the class of nodes that are typically not reached by WTM contagions initialized with cluster seeding—are all relatively isolated, so one can construe them as peripheral nodes in the network [2]. Interestingly, our experimental results suggest that the inability to reach a node [i.e., nodes in classes (3) and (4)] is related to a global network property (i.e., whether it is “isolated”), whereas the inability to seed a large contagion [i.e., nodes in classes (2) and (4)] depends on both local and global network properties.

In the bottom row of Supplementary Fig. 4, we show properties of the metro proximities  $\{\psi_i\}$  for the London transit network. In panel (d), we show a histogram of the frequencies of nodes at a given metro proximity  $\psi_i$ , and we note that most nodes are 5–20 edges from a metro station. In panel (e), we give a scatter plot of the nodes’ total degrees  $\{d_i\}$  versus their metro proximities  $\{\psi_i\}$ . Note that the metro stations (for which  $\psi_i = 0$ ) have large degrees relative to the other nodes: their mean degree is 5, whereas the mean degree of all nodes is approximately 2.59. In panel (f), we show that isolated nodes, which by definition are distant from metro stations, also tend to be distant to other nodes in the London transit network. Specifically, for each node  $i$ , we plot the mean length of the shortest path from it to the remaining nodes  $j \in \mathcal{V} \setminus \{i\}$  versus its metro proximity  $\psi_i$ . Nodes with large values of  $\psi_i$  are also more distant (on average) to the other nodes. It is therefore appropriate to use the term “isolated” to describe these nodes.

Combining the results from Supplementary Figs. 3 and 4, we find when we ignore the activation times of infinity that WTM maps have larger values of  $\rho$  when  $T$  is in the approximate interval  $(0.1, 0.2)$  than when  $T$  takes other values. The activation times of infinity result from the existence of a few nodes  $i$  such that WTM contagions that we initialize with cluster seeding centered at those nodes tend not to spread very far. These nodes tend to be in class (2), and they are often either 2–3 edges from metro stations or are isolated nodes. Finally, when  $T$  is sufficiently large so that nodes belong to class (3) (e.g., as occurs for  $T > 0.2$ ), then the values of  $\rho$  are comparatively very small. Recall that nodes in class (3), which almost never adopt contagions, are relatively isolated nodes in the network.

## “Egocentric” Analysis of Geometry

Thus far, we have studied geometry through the Pearson correlation coefficient  $\rho$  given by Eq. (18). As we discussed in Sec. III D of the main text (and also see Supplementary Note 7),  $\rho$  describes the correlation between node-to-node distances  $\{m(i, j)\}$  for the intrinsic locations  $\{\mathbf{w}^{(i)}\}$  [see Eq. (16)] and node-to-node distances  $\{m^{(\text{WTM})}(i, j)\}$  for the point clouds  $\{\mathbf{x}^{(i)}\}$ ,  $\{\mathbf{y}^{(i)}\}$ , or  $\{\mathbf{z}^{(i)}\}$  that result from a WTM map [see Eq. (17)]. We calculate the correlation  $\rho$  using the  $N(N - 1)/2$  unordered pairs of nodes  $(i, j) \in \mathcal{V} \times \mathcal{V}$  (where  $i \neq j$ ), and one can interpret it as comparing the geometry of these two point clouds at a “network level.” To gain further insight, we now compare the geometry of the two point clouds at a “node level” by computing “egocentric” correlation coefficients that consider only node-to-node distances that involve a particular node  $i$ . Specifically, we study a set of Pearson correlation coefficients  $\{\hat{\rho}_i(T)\}$  for a given  $i \in \mathcal{V}$ .

We introduce the egocentric correlation coefficient  $\hat{\rho}_i(T)$  for the regular WTM map  $\mathcal{V}_{sto}\{\mathbf{x}^{(i)}\}$ , and we note that one can apply it to any version of a WTM map. For each node  $i$ , we study the Pearson correlation coefficient  $\hat{\rho}_i(T)$  that relates node-to-node distances  $\{m(i, j)\}$  from node  $i$  to all nodes  $j \in \mathcal{V}$  with respect to the intrinsic locations  $\{\mathbf{w}^{(i)}\}$  [see Eq. (16)] to the node-to-node distances  $\{m^{(\text{WTM})}(i, j)\}$  from node  $i$  to all nodes  $\{j\} \in \mathcal{V}$  for a point cloud  $\{\mathbf{x}^{(i)}\}$  that results from a WTM map [see Eq. (17)]. Specifically, we



compute

$$\hat{\rho}_i(T) = \frac{\sum_{j=1}^N [m(i, j) - \overline{m(i, j)}] [m^{(\text{WTM})}(i, j) - \overline{m^{(\text{WTM})}(i, j)}]}{\sqrt{\sum_{j=1}^N [m(i, j) - \overline{m(i, j)}]^2} \sqrt{\sum_{j=1}^N [m^{(\text{WTM})}(i, j) - \overline{m^{(\text{WTM})}(i, j)}]^2}}, \quad (1)$$

where the bar above a variable indicates that we are taking its mean for all nodes  $j \in \mathcal{V}$ . Note the strong similarity between Eq. (1) and Eq. (18); the only difference is that the summations in Eq. (1) are over  $j$  rather than over both  $j$  and  $i$ .

In Supplementary Fig. 5, we study egocentric correlation coefficients  $\{\hat{\rho}_i(T)\}$  for WTM maps on the London transit network for two values of the threshold  $T$ . In the top panels, we show results for the map  $\mathcal{V}_{sto}\{\mathbf{x}^{(i)}\}$ ; in the bottom panels, we show results for the map  $\mathcal{V}_{sto}\{\mathbf{y}^{(i)}\}$ . For both maps, we handle the activation times of infinity by neglecting them. In the left column, we plot egocentric Pearson correlation coefficients  $\{\hat{\rho}_i(T)\}$  versus the metro proximities  $\{\psi_i\}$  for the threshold values  $T \in \{0.1, 0.18\}$ . We show values only for nodes that are not in class (2). Note that the larger value of  $T$  tends to have larger values of  $\hat{\rho}_i(T)$  in both panels (a) and (d). We highlight this feature further in the center column by plotting  $[\hat{\rho}_i(0.18) - \hat{\rho}_i(0.1)]$  versus  $\psi_i$ . The solid and dashed curves, respectively, indicate the mean values for  $T = 0.1$  and  $T = 0.18$  for a given  $\psi_i$ . In the right column of Supplementary Fig. 5, we plot histograms of the frequencies of observed values  $[\hat{\rho}_i(0.18) - \hat{\rho}_i(0.1)]$ , and we note that they appear to have heavy tails:  $[\hat{\rho}_i(0.18) - \hat{\rho}_i(0.1)]$  tends to be small and positive for most nodes  $j \in \mathcal{V}$ , but there are some nodes for which  $[\hat{\rho}_i(0.18) - \hat{\rho}_i(0.1)]$  is rather large.

## Summary of Experiments with the London Transit Network

We studied WTM contagions on a London transit network in which nodes are intersections that are connected either by roads (which we interpreted as geometric edges) or by metro lines (which we interpreted as non-geometric edges). Similar to our study of WTM contagions on synthetic networks, we found that WFP and ANC arise for WTM contagions on this empirical network, and the type of epidemic propagation depends significantly on the contagion threshold  $T$ . We studied these WFP and ANC by analyzing the geometry of WTM maps, and we observed that the geometry of point clouds that result from WTM maps agree better with the geometry of the nodes' intrinsic locations on Earth's surface for values of  $T$  in the approximate range (0.1, 0.2) than for other values of  $T$ . To obtain this result, we examined situations with activation times of infinity in two different ways: (1) setting those times to be  $2N$ , as in the synthetic examples in the main text; and (2) ignoring these values in our subsequent calculations. We found the latter approach to be more useful for the London transit network. Our investigation led us to assign nodes into four classes based on their ability to initiate large contagions and consistently adopt contagions, and our calculations yielded an interesting connection between the proximity of nodes to metro stations and their behavior with respect to WTM contagions.

## Supplementary Note 2 Denoising Networks with WTM Maps

The embedding of a network into a metric space has numerous applications, ranging from the control and optimization of dynamics to network “denoising” (i.e., the identification of spurious and missing edges). In this section, we highlight one application of WTM maps: the identification of noisy edges.

Our methodology for denoising proceeds as follows. Given the WTM map for a network, we determine the length  $m^{(\text{WTM})}(i, j)$  [given by Eq. (17) in Supplementary Note 7] in the embedding space of each edge  $(i, j) \in \mathcal{E}$ . Because we expect non-geometric edges to have larger lengths than geometric edges, examining the set of edge lengths  $\{m^{(\text{WTM})}(i, j)\}_{(i, j) \in \mathcal{E}}$  allows one to infer edge type. For example, by studying the distribution of edge lengths, one can choose a *partitioning threshold* to partition the edges into classes (i.e., geometric and non-geometric) by comparing their lengths to the partitioning threshold. There exist various heuristic approaches for selecting such a partitioning threshold, so we will consider all possible partitioning thresholds in our experiments. To do this, we construct *receiver operating characteristic* (ROC) curves that examine the fraction of false positives and false negatives as the partitioning threshold is increased from the smallest edge length to the largest edge length.

To gauge the performance of this approach for denoising networks, we compare our results to a popular approach based on subgraph statistics. For each edge  $(i, j) \in \mathcal{E}$ , we compute the Jaccard index  $|\mathcal{N}_i \cap \mathcal{N}_j|/|\mathcal{N}_i \cup \mathcal{N}_j|$  to measure the overlap of the set  $\mathcal{N}_i \triangleq \{k \in \mathcal{V} : A_{ik} \neq 0\}$  of nodes that neighbor node  $i$  with the set  $\mathcal{N}_j \triangleq \{k \in \mathcal{V} : A_{jk} \neq 0\}$  of nodes that neighbor node  $j$  [4]. Similar to our approach of comparing the edge lengths to some partitioning threshold, one can compare the edges’ Jaccard indices to a partitioning threshold and then vary the partitioning threshold to yield a ROC curve. This allows for a direct comparison between the two approaches.

Note that the approach of Ref. [4] is “local”—i.e., each edge is classified based on the properties of the subgraph that contains nodes and edges that are adjacent to that edge—whereas denoising based on WTM maps is a “global” approach that uses an entire network for the denoising procedure.

### Denoising the London Transit Network

In our first experiment, we examine the utility of WTM maps for identifying the metropolitan lines in the London transit system that we study in Sec. I F of the main text. Because this noisy geometric network results from the merging of two network layers—a road network and the metropolitan system—our aim in this context is to disaggregate the two network layers based on the assumption that metropolitan lines connect nodes that are farther apart than those that are connected by roads. In this experiment, we purposely do not utilize the known node locations, as we are interested in the ability of WTM maps to identify the metro lines based on the network structure alone.

In Supplementary Fig. 6(a), we plot ROC curves for symmetric WTM maps that we construct with various choices of the WTM threshold  $T$ . For these maps, we set the activation times of infinity to  $2N$ . Perfect inference of the noisy edges would correspond to a ROC curve in which the true-positive rate is always 1 for any nonzero false-positive rate. We also note that the ROC curves for the WTM depend strongly on  $T$ . For  $T = 0.1$ , the curve shows poor performance, similar to what one obtains using a Jaccard-index approach [4]. For larger  $T$  (i.e.,  $T \gtrsim 0.1$ ), denoising based on WTM maps outperforms this other approach. Recall for the London transit network that when  $T$  surpasses 0.1, we observe an increase in WFP [which is indicated by the larger values of  $\rho$  in Supplementary Fig. 3(c) when  $T$  surpasses 0.1]. The fact the ROC curves are still very high in Supplementary Fig. 6(a) when  $T > 0.2$  is somewhat unexpected, because we previously observed that there is disagreement between the geometry of WTM maps and the geometry

of the actual London transit network for this range [see the drop in  $\rho$  values that occurs in Supplementary Fig. 3(c) as  $T$  surpasses 0.2]. Interestingly, these results imply that the length of edges in the WTM map can still be very predictive for classifying edges as geometric or non-geometric even when the geometry of the WTM map disagrees with that of the actual network.

Here, we perform additional experiments to explore network properties that can help to shed light on these results.

## Denoising Noisy Two-Dimensional Square Lattices

We now do an experiment to highlight that local algorithms based on an assumption about the local network structure—e.g., a prevalence of 3-cycles (i.e., triangles) [4]—can be very inaccurate if that assumption is invalid. In particular, modern road networks are known to exhibit a prevalence of subgraphs other than 3-cycles [5] (e.g., city blocks can give rise to 4-cycles), and we therefore expect the lack of 3-cycles to be a significant factor that influences the results in Supplementary Fig. 6(a).

In our experiment, we examine the inference of noisy edges for a synthetic noisy geometric network with  $N = 40^2$  nodes that are embedded as a 2D lattice with periodic boundary conditions. To construct the “substrate” geometric network, we place  $d^{(G)} = 4$  geometric edges for each node to connect nearest-neighbor nodes both horizontally and vertically. We then add  $40^2$  non-geometric (i.e., “noisy” edges) uniformly at random to pairs of nodes that are not already connected by a geometric edge. Therefore, each node has  $d^{(NG)} = 2$  non-geometric edges on average. We note that this procedure—which adds non-geometric edges to a network that already has geometric edges—is identical to the procedure that we used for families **(b)** and **(d)** of the noisy ring networks in Supplementary Note 3.

In Supplementary Fig. 6(b), we plot ROC curves for the inference of noisy edges via symmetric WTM maps using various values of the WTM threshold  $T$ . As before, we set the activation times of infinity to  $2N$ . Note that the best ROC curve corresponds to  $T = 0.3$ , and that the ROC curves for WTM maps with  $T \in \{0.2, 0.3, 0.4\}$  are much better than that for the Jaccard-index approach.

To more precisely compare the different ROC curves for different  $T$ , in Supplementary Fig. 6(c), we plot the *area under the ROC curve* (AUC) as a function of  $T$  (dashed curve). We use crosses to indicate the values of  $T$  that we used to generate Supplementary Fig. 6(b). To gauge the performance of inference using WTM maps, we note that the best attainable AUC is a value of 1 (which is almost reached for the WTM map with  $T = 0.3$ ). Using the horizontal red line, we show the AUC for the Jaccard-index approach. Its value is approximately 0.5, indicating that it is comparable to random guessing in this scenario.

## Denoising Noisy Ring Lattices with Removal of Geometric Edges

In our final experiment, we examine the effect of stochasticity on the inference of noisy edges. In particular, we explore the inference of noisy geometric networks in which we have removed some percentage of the geometric edges. We consider family **(a)** (see Supplementary Note 3) of the noisy ring lattices, in which nodes are evenly placed on the unit circle in  $\mathbb{R}^2$ . We construct networks with  $N = 200$  nodes, where each node has  $d^{(NG)} = 1$  non-geometric edge and we consider various choices of geometric degree  $d^{(G)}$ . We then remove some percentage of the  $Nd^{(G)}/2$  geometric edges—chosen uniformly at random—to include stochasticity. A nice feature of this experiment is that we can simultaneously increase  $d^{(G)}$  and increase the edge-removal percentage so that the expected number of geometric edges (after removals) remains constant. In this procedure, note that although the number of geometric edges after removal is constant by construction, the mean length of the geometric edges tends to increase as we consider higher

levels of stochasticity (i.e., by adding and then removing a larger number of edges).

In Supplementary Fig. 7, we plot ROC curves for the inference of noisy edges using symmetric WTM maps in which we set the activation times of infinity to  $2N$ . In panels (a)–(d), we show results for four networks, which we construct using progressively larger values of  $d^{(G)}$  and in which we remove an associated larger percentage of geometric edges so that, on average, every node has  $d^{(G)} = 20$  geometric edges after the removals. In each panel, we depict ROC curves for several values of the WTM threshold  $T$  as well as for the Jaccard-index approach. Note that the ROC curves generally become lower as one moves from panel (a) to (d). In panel (a), for example, WTM maps for all  $T$  values and the Jaccard-index approach lead to the accurate inference of the noisy edges. In panel (d), however, we find that the WTM map with  $T = 0.2$  leads to the best ROC curve. Our main finding is that incorporating stochasticity into the presence of geometric edges inhibits the successful inference of noisy edges. Depending on the value of  $T$  and the network parameters, denoising based on WTM maps can perform either better or worse than a local approach based on the Jaccard index.

## Summary of Experiments for Denoising Networks

Our experiments highlight the use of WTM maps for the denoising of networks. We now briefly discuss the advantages and drawbacks of this novel technique in comparison to other approaches; an in-depth exploration would be very interesting, but it is well beyond the scope of the present paper. One class of previous approaches are “local” approaches that make an assumption about local network properties, such as a prevalence of 3-cycles (i.e., triangles), and then infer “noisy” edges to be the ones that do not follow this assumption. One can attempt to infer whether a particular edge is consistent with such an assumption by examining a Jaccard index or another subgraph statistic [4, 6, 7]. Because these are “local” approaches, they have the advantage of being fast and straightforward to compute. In contrast, our approach based on WTM maps is reminiscent of “global” approaches that leverage a model for an entire network (i.e., as opposed to a model for the local subgraph structure) to find edges that do not adhere to the model [8–10]. We note that these prior efforts often have focused on the problem of identifying missing (rather than spurious) edges, although these problems are closely related [9].

In our experiments, we have illustrated examples of noisy geometric networks in which a global approach based on WTM maps can be advantageous to a local approach. We demonstrated that the global perspective of the WTM can be beneficial for denoising networks that fail to have a sufficient prevalence of 3-cycles (so that methods based on, e.g., the Jaccard index, do not perform well in those scenarios). We have demonstrated this situation both for the London transit network and for noisy 2D square lattices. Furthermore, even in scenarios in which 3-cycles are prevalent, we found that the WTM and Jaccard-index approaches show similar levels of performance [see Supplementary Fig. 7(a)]. For noisy ring lattices that also include stochasticity in the geometric edges, we found (depending on the value of the WTM threshold  $T$ ) that an approach based on WTM maps can lead to either higher or lower AUC values than an approach based on the Jaccard index.

### Supplementary Note 3 Generalizations of the Noisy Ring Lattice

In the main text, we analyzed the WTM on noisy ring lattices. In this section, we review our construction of noisy ring lattices and introduce three additional families of noisy geometric networks that use an underlying ring manifold. In these families, we introduce heterogeneity into the nodes' geometric and non-geometric degrees, which we now denote, respectively, by  $d_i^{(G)}$  and  $d_i^{(NG)}$  for a given node  $i$ . We denote their means over the nodes by  $\langle d_i^{(G)} \rangle = N^{-1} \sum_i d_i^{(G)}$  and  $\langle d_i^{(NG)} \rangle = N^{-1} \sum_i d_i^{(NG)}$ , respectively. We therefore adjust our definition of the ratio  $\alpha$  to denote the ratio of the mean non-geometric degree to the mean geometric degree:

$$\alpha = \langle d_i^{(NG)} \rangle / \langle d_i^{(G)} \rangle. \quad (2)$$

We note that it is equivalent to state that  $\alpha$  denotes the ratio of the number of non-geometric edges to the number of geometric edges in a given network.

#### Families of Noisy Geometric Networks on a Ring Manifold

We now define four families of noisy geometric networks on a ring manifold given by the unit circle in  $\mathbb{R}^2$ . We label these families as **(a)**, **(b)**, **(c)**, and **(d)**. In Supplementary Fig. 8, we illustrate an example network for each family and plot its corresponding adjacency matrix and degree distribution.

- **Family (a).** To generate the noisy ring lattice that we studied in the main text, we place  $N$  nodes evenly on the unit circle in  $\mathbb{R}^2$  so that each node  $i$  has location  $\mathbf{w}^{(i)} = [\cos(\theta_i), \sin(\theta_i)]^T$  with  $\theta_i = 2\pi i/N$ . We then add geometric edges between neighboring node pairs  $(i, j) \in \mathcal{V} \times \mathcal{V}$ , so that each node  $i$  has exactly  $d_i^{(G)} = \langle d_i^{(G)} \rangle$  geometric edges. That is, we connect each node to its nearest  $d_i^{(G)}/2$  neighbors on each side, and we note that  $d_i^{(G)}$  is even because of symmetry. We then assign non-geometric edges randomly using (a slight modification of) the configuration model [11] so that each node has exactly  $d_i^{(NG)} = \langle d_i^{(NG)} \rangle$  non-geometric edges. As in the configuration model, we connect ends of edges (i.e., “stubs”) to each other uniformly at random, but we disallow self-edges and multi-edges. Our implementation of the configuration model is a slight modification of the original version, because we want to guarantee that the set of geometric edges is disjoint from the set of non-geometric edges. Specifically, if we propose a candidate edge between two nodes that would lead to a disallowed situation (i.e., it would lead to a self-edge, multi-edge, or an edge that is already a geometric edge), then we discard the candidate edge, and we propose a new candidate edge as prescribed by the configuration model. The resulting network is a  $(\langle d_i^{(G)} \rangle + \langle d_i^{(NG)} \rangle)$ -regular network that contains  $N\langle d_i^{(G)} \rangle/2$  geometric edges and  $N\langle d_i^{(NG)} \rangle/2$  non-geometric edges. The geometric edges form a deterministic backbone (as in the Newman-Watts variant [12, 13] of the Watts-Strogatz model [14]), whereas we obtain non-geometric edges through a stochastic process.
- **Family (b).** Our first generalization of the noisy ring lattice in family **(a)** is to allow heterogeneity in the number of non-geometric edges that are incident to a given node  $i$  (i.e., its non-geometric degree  $d_i^{(NG)}$ ). The total number of non-geometric edges is still equal to the constant  $N\langle d_i^{(NG)} \rangle/2$ , but we now distribute them uniformly at random among the  $\frac{N \cdot (N-1-d^{(G)})}{2}$  possible edge locations that are unoccupied by geometric edges. Hence, the subgraph that consists only of non-geometric edges limits to an Erdős-Rényi (ER) network when  $N \gg d^{(G)}$  [11]. The distribution of non-geometric degrees is thus a binomial distribution that is centered at  $\langle d_i^{(NG)} \rangle$ .
- **Family (c).** Our second generalization of **(a)** is to allow heterogeneity in the node locations on the unit circle in  $\mathbb{R}^2$ . Constraining geometric edges by distance, in turn, leads to heterogeneity in the number of geometric edges

that are incident to a given node  $i$  (and hence in its geometric degree  $d_i^{(G)}$ ). To make such a generalization in a tunable manner, we assign the node locations (or, equivalently, the angles  $\{\theta_i\}$  in the case of the unit circle) to be evenly spaced as for family **(a)**, and we then perturb these locations using a random variable  $\delta\theta_i$ , so that the location for each node  $i$  is given by  $[\cos(\theta_i + \delta\theta_i), \sin(\theta_i + \delta\theta_i)]^T$ . We consider a Gaussian-distributed random variable  $\delta\theta_i \sim \mathcal{N}(0, (s\frac{2\pi}{N})^2)$ , where one can vary  $s$  to adjust the amount of heterogeneity in node location along a ring manifold. The choice  $s = 0$  recovers the original node locations, and  $s \rightarrow \infty$  corresponds to sampling locations on the unit circle uniformly at random. Unless we specify otherwise, we use  $s = 1/2$ . To generate geometric edges, we choose a parameter  $\epsilon > 0$  and place edges between all pairs of nodes  $i$  and  $j$  such that  $|\theta_i - \theta_j| < \epsilon$ . To compare networks from family **(c)** to networks from families **(a)** and **(b)**, for which the nodes have the identical geometric degree  $d_i^{(G)} = \langle d_i^{(G)} \rangle$ , we choose the parameter  $\epsilon$  so that each network in family **(c)** has exactly  $N\langle d_i^{(G)} \rangle/2$  edges.

- Family **(d)**. Our final network family combines the generalizations from families **(b)** and **(c)** so that there is heterogeneity in the non-geometric degrees  $\{d_i^{(NG)}\}$  (where the mean is  $\langle d_i^{(NG)} \rangle$ ), the geometric degrees  $\{d_i^{(G)}\}$  (where the mean is  $\langle d_i^{(G)} \rangle$ ), and the node locations  $\{\mathbf{w}^{(i)}\}$ .

## Perturbed Bifurcation Results

Equations (1) and (2) in the main text give sequences of critical thresholds that determine WFP and ANC for large networks of family **(a)**. Recall that the degrees  $d_i^{(G)}$  and  $d_i^{(NG)}$  for each node  $i$  are deterministic and constant for family **(a)**. However, here we introduce various types of stochasticity (and hence heterogeneity) for these degrees in network families **(b)**–**(d)**. Because of such heterogeneity, the critical thresholds that we derived previously for network family **(a)** no longer accurately describe the WTM contagion dynamics. However, based on numerical experiments, we find that Eqs. (1) and (2) in the main text still describe contagion dynamics at a given node  $i$  if we use the correct geometric and non-geometric degrees. Specifically, the ability of node  $i$  to adopt a contagion via WFP when it has no infected non-geometric neighbors is given approximately by Eq. (13) with the substitutions  $d^{(G)} \text{stod}_i^{(G)}$  and  $d^{(NG)} \text{stod}_i^{(NG)}$ . Similarly, the ability of node  $i$  to adopt a contagion via ANC is given approximately by Eq. (7) with the substitutions  $d^{(G)} \text{stod}_i^{(G)}$  and  $d^{(NG)} \text{stod}_i^{(NG)}$ . Hence, for each node  $i \in \mathcal{V}$ , there are sequences of critical thresholds,  $\{T_k^{(\text{WFP})}\}$  and  $\{T_k^{(\text{ANC})}\}$ , that are (potentially) specific to that node. Consequently, the nodes can exhibit qualitatively dissimilar contagion dynamics with respect to WFP and ANC. For example, for a given threshold  $T$ , some nodes can have geometric and non-geometric degrees that support WFP but no ANC, whereas other nodes can have degrees that support both WFP and ANC. Nevertheless, one can construe the bifurcation analysis that we developed for family **(a)** as an approximate bifurcation analysis for the other families. In this light, note that if the degree heterogeneities are sufficiently small compared to the mean degrees, then we still identify four different qualitative regimes of WTM contagion dynamics that are marked by the absence versus presence of WFP and ANC. However, the boundaries that separate these regimes are perturbations of what we found for family **(a)**.

More precisely, for each node  $i$ , let  $\delta_i^{(G)} = d_i^{(G)} - \langle d_i^{(G)} \rangle$  denote the difference between its geometric degree and the mean geometric degree. Similarly, let  $\delta_i^{(NG)} = d_i^{(NG)} - \langle d_i^{(NG)} \rangle$  denote the difference between its non-geometric degree and the mean non-geometric degree. Restricting our attention to the critical thresholds given by Eqs. (1) and (2) in the main text for  $k = 0$  (although one can write similar

expressions for other values of  $k$ ), we can express the critical thresholds in terms of  $\delta_i^{(G)}$  and  $\delta_i^{(NG)}$  as

$$T_0^{(ANC)}(\delta_i^{(G)}, \delta_i^{(NG)}) \triangleq \frac{1 + \frac{\delta_i^{(NG)}}{\langle d_i^{(NG)} \rangle}}{\alpha^{-1} + 1 + \frac{(\delta_i^{(G)} + \delta_i^{(NG)})}{\langle d_i^{(NG)} \rangle}} \quad (3)$$

$$T_0^{(WFP)}(\delta_i^{(G)}, \delta_i^{(NG)}) \triangleq \frac{1 + \frac{\delta_i^{(G)}}{\langle d_i^{(G)} \rangle}}{2 + 2\alpha + \frac{2(\delta_i^{(G)} + \delta_i^{(NG)})}{\langle d_i^{(G)} \rangle}}. \quad (4)$$

Expressions (3) and (4) summarize the effect of degree heterogeneity on the critical thresholds that describe a WTM contagion at a given node  $i$ . When there is no degree heterogeneity (e.g., family **(a)**, where  $\delta_i^{(G)} = \delta_i^{(NG)} = 0$ ), we recover our results from the main text:  $T_0^{(ANC)}(0, 0) = 1/(\alpha^{-1} + 1)$  and  $T_0^{(WFP)}(0, 0) = 1/(2 + 2\alpha)$ . Meanwhile, when  $\delta_i^{(G)}$  and  $\delta_i^{(NG)}$  are both small compared to their respective means  $\langle d_i^{(G)} \rangle$  and  $\langle d_i^{(NG)} \rangle$ , the perturbed critical thresholds are approximately equal to those for the mean degree. In other words,  $T_0^{(ANC)}(\delta_i^{(G)}, \delta_i^{(NG)}) \approx T_0^{(ANC)}(0, 0)$  when  $(\delta_i^{(G)} + \delta_i^{(NG)})/\langle d_i^{(NG)} \rangle$  is small, and  $T_0^{(WFP)}(\delta_i^{(G)}, \delta_i^{(NG)}) \approx T_0^{(WFP)}(0, 0)$  when  $(\delta_i^{(G)} + \delta_i^{(NG)})/\langle d_i^{(G)} \rangle$  is small. We therefore interpret our bifurcation analysis for network family **(a)** as an approximate bifurcation analysis for network families **(b)**–**(d)**. We expect our interpretation to be increasingly accurate as the mean degrees become larger relative to the heterogeneity in the degrees.

In Supplementary Fig. 9, we plot curves that indicate the node-specific critical thresholds given by Eqs. (3) and (4) for the  $(T, \alpha)$  parameter plane. We show results for  $d^{(G)} = 20$  and  $d^{(NG)} \in [0, 20]$  for several choices of  $\delta_i^{(G)}$  and  $\delta_i^{(NG)}$  such that  $|\delta_i^{(G)}| \leq 2$  and  $|\delta_i^{(NG)}| \leq 2$ . Panels (a)–(d), respectively, demonstrate the heterogeneity that arises in  $\delta_i^{(G)}$  and  $\delta_i^{(NG)}$  from network families **(a)**–**(d)**. Note for all panels that there exist parameter regimes in which the nodes support similar contagion phenomena (i.e., WFP and no ANC, WFP and ANC, no WFP and ANC, or no WFP and no ANC) even through their degrees are heterogeneous. Therefore, one can construe the set of curves that one obtains for multiple values of  $\delta_i^{(G)}$  and  $\delta_i^{(NG)}$  as a “thickening” of the boundary between regions of qualitatively different dynamics. In other words, as we vary parameters, we see that the transitions between regions of different dynamics can occur for slightly different parameter values for different nodes in a network. Note, however, that this interpretation does not take into account the distribution of node degrees, as we have only shown the critical threshold curves in Supplementary Fig. 9 for degrees that are near the mean degrees (i.e., for  $|\delta_i^{(G)}| \leq 2$  and  $|\delta_i^{(NG)}| \leq 2$ ). As we discussed in this Supplementary Note, families **(b)**–**(d)** lead to networks with heterogeneous node degrees, and it is then possible that  $|\delta_i^{(G)}| > 2$ . (For example, the non-geometric degrees follow a binomial distribution for families **(b)** and **(d)**.) Therefore, one can construe our bifurcation analysis for family **(a)** as an approximate bifurcation analysis for families **(b)**–**(d)** only when the heterogeneities in the two types of degrees are both sufficiently small when compared to the means  $\langle d_i^{(G)} \rangle$  and  $\langle d_i^{(NG)} \rangle$ .

## Supplementary Note 4 A Set of Filtrations Defines a Metric

In this section, we prove that the set of activation times for a WTM contagion with threshold  $T$  on a network induces a metric on the node set  $\mathcal{V} = \{1, 2, \dots, N\}$ . Let  $\tilde{x}_j^{(i)}$  denote the activation time (which we assume to be finite for all node pairs  $(i, j) \in \mathcal{V} \times \mathcal{V}$ ) for node  $i$  for a contagion initialized with the seed node  $j$ . We will show that  $(\mathcal{V}, m^{(\text{WTM})}(i, j))$  is a discrete metric space with metric  $m^{(\text{WTM})}(i, j) = \tilde{x}_j^{(i)} + \tilde{x}_i^{(j)}$ . However, rather than showing this result for the specific case of a WTM contagion, we will prove a more general result using the observation that the growing set of infected nodes during one realization of a WTM contagion defines a “filtration” of the node set  $\mathcal{V}$ . We will therefore prove that any “complete” and “consistent” set of filtrations (see the definitions below) on a finite set  $\mathcal{V}$  induces a metric on  $\mathcal{V}$ . Subsequently, because the filtrations that result from realizations of a WTM contagion on a given network with contagion seeds  $\{j\}$  for  $j \in \{1, \dots, N\}$  satisfy the conditions of completeness and consistency, it follows that  $m^{(\text{WTM})}(i, j) = \tilde{x}_j^{(i)} + \tilde{x}_i^{(j)}$  is a metric on  $\mathcal{V}$  whenever  $\tilde{x}_i^{(j)} < \infty$  for every  $i, j \in \mathcal{V}$ .

### Complete and Consistent Filtrations

Before proving that the set of activation times—and, more generally, any “complete” and “consistent” set of “filtrations”—leads to a metric, we give a few definitions.

**Definition: Filtration.**

Consider a sequence of sets  $\mathcal{N}_t$  for  $t \in \{1, 2, \dots\}$ . The sequence of sets is called a *filtration* [15–17] if it has the property that  $\mathcal{N}_t \subseteq \mathcal{N}_{t+1}$  for all  $t$ <sup>1</sup>.

**Definition: Completeness.**

Let  $\mathcal{V}$  be a finite set with cardinality  $|\mathcal{V}| = N$ . We define a set of filtrations to be *complete* if there are  $N$  filtrations of the following form: for every  $j \in \mathcal{V}$ , there exists a filtration such that

$$\{j\} = \mathcal{N}_0(j) \subseteq \mathcal{N}_1(j) \subseteq \mathcal{N}_2(j) \subseteq \dots \subseteq \mathcal{N}_{t_j^*}(j) = \mathcal{V}. \quad (5)$$

Note that the filtration  $\{\mathcal{N}_t(j)\}$  consists of nested sets  $\mathcal{N}_t(j)$ , where the innermost set is the element  $\{j\}$  and the outermost set (i.e., the  $t_j^*$ th set) is the complete set  $\mathcal{V}$  of indices.

**Definition: Consistency.**

Let  $\mathcal{V}$  be a finite set with cardinality  $|\mathcal{V}| = N$ , and consider a set of filtrations in which the  $j$ th filtration corresponds to nested sets  $\{\mathcal{N}_t(j)\}$  that are indexed by  $t$ . We define the set of filtrations to be *consistent* if, for any two filtrations  $\{\mathcal{N}_t(i)\}$  and  $\{\mathcal{N}_t(j)\}$  from the set, the following is true:

$$\mathcal{N}_t(i) \subseteq \mathcal{N}_\tau(j) \implies \mathcal{N}_{t+1}(i) \subseteq \mathcal{N}_{\tau+1}(j), \quad (6)$$

where the indices  $t$  and  $\tau$  can be different from each other.

---

<sup>1</sup>One can also define a filtration sequence of filtrations using the superset relation  $\mathcal{N}_t \supseteq \mathcal{N}_{t+1}$ . Additionally, one can also generalize the notion of a filtration to concepts like “zigzag filtrations,” which allow both subset and superset relations amidst the sequence of sets [18].



## Filtration-Induced Metrics

**Theorem:** *A Metric Induced by Filtrations.*

Let  $\mathcal{V}$  be a finite set with cardinality  $|\mathcal{V}| = N$ , and let  $\mathcal{N}_t(j)$  denote sets that define a complete and consistent set of filtrations on  $\mathcal{V}$ . Additionally, let  $t_j^{(i)}$  denote the smallest index  $t$  such that  $i \in \mathcal{N}_t(j)$ . It then follows that  $m(i, j) = t_i^{(j)} + t_j^{(i)}$  defines a metric on the set  $\mathcal{V}$ .

*Proof:* First, we show that  $m(i, j) \geq 0$  and that  $m(i, j) = 0$  implies  $i = j$ . Note that  $t_j^{(i)} \geq 0$  for any  $i, j \in \mathcal{V}$ ; this implies that  $m(i, j) \geq 0$ . Similarly,  $m(i, j) = 0$  requires  $t_j^{(i)} = t_i^{(j)} = 0$ , which in turn requires that  $\mathcal{N}_0(j) = \{i\}$  (and  $\mathcal{N}_0(i) = \{j\}$ ). However, we know by definition that  $\{j\} = \mathcal{N}_0(j)$  (and  $\{i\} = \mathcal{N}_0(i)$ ), so it must be the case that  $i = j$ . It is trivial to show that  $m(i, j) = m(j, i)$ . Finally, we complete the proof by showing that  $m(i, j)$  satisfies the triangle inequality:  $m(i, j) \leq m(i, k) + m(k, j)$ . This step is a bit more complicated, and it relies on the completeness and consistency of the set of filtrations. Using the definition of  $m(i, j)$ , it suffices to show that  $t_j^{(i)} \leq t_j^{(k)} + t_k^{(i)}$ . Using the notation  $a = t_j^{(i)}$ ,  $b = t_j^{(k)}$ , and  $c = t_k^{(i)}$ , we will prove that  $a \leq b + c$ . Because the result is trivial when  $b \geq a$  due to the non-negativity of  $c$ , we can assume that  $a > b$ . By definition, it must be the case that  $i \in \mathcal{N}_a(j)$ ,  $k \in \mathcal{N}_b(j)$ , and  $i \in \mathcal{N}_c(k)$ . Because  $b < a$ , it must also be the case that  $\mathcal{N}_b(j) \subseteq \mathcal{N}_a(j)$ . We now consider  $\{k\} = \mathcal{N}_0(k) \subseteq \mathcal{N}_b(j)$ , which uses the completeness of the filtrations. Using the consistency property, it follows that  $\mathcal{N}_1(k) \subseteq \mathcal{N}_{b+1}(j)$ ,  $\mathcal{N}_2(k) \subseteq \mathcal{N}_{b+2}(j)$ , and so on. Repeating this procedure demonstrates that  $\mathcal{N}_c(k) \subseteq \mathcal{N}_{b+c}(j)$ . Noting that  $i \in \mathcal{N}_c(k)$ , it follows that  $i \in \mathcal{N}_{b+c}(j)$ . It follows, in turn, that  $t_j^{(i)} \leq b + c$ , which is equivalent to  $a \leq b + c$ .  $\square$

**Corollary:** *A Metric Induced by WTM Contagions.*

Consider a network with node set  $\mathcal{V}$  and edge set  $\mathcal{E}$  that consists of a single connected component. Let  $\tilde{x}_j^{(i)}$  denote the activation time of node  $i$  for a WTM contagion with seed  $\{j\}$ . As before, we assume that  $\tilde{x}_j^{(i)}$  is finite for all node pairs  $(i, j) \in \mathcal{V} \times \mathcal{V}$ . It then follows that  $m(i, j) = \tilde{x}_i^{(j)} + \tilde{x}_j^{(i)}$  is a metric on the node set  $\mathcal{V}$ .

*Proof:* It suffices to show that  $N$  realizations of a WTM contagion with the set of contagion seeds  $\mathcal{S}^{(j)} = \{j\}$  (for  $j \in \mathcal{V}$ ) produces a complete and consistent set of filtrations on  $\mathcal{V}$ . It will be convenient to use the notation  $t_j^{(i)} = \tilde{x}_j^{(i)}$ . We first prove completeness. Let  $\mathcal{N}_t(j)$  denote the set of nodes for realization  $j$  that have adopted the contagion by time  $t$ . Note that  $\mathcal{N}_0(j) = \mathcal{S}^{(j)} = \{j\}$  for each  $j$ . Additionally,  $\mathcal{N}_t(j) \subseteq \mathcal{N}_{t+1}(j)$  for any  $t$ , as nodes cannot unadopt a contagion during a time step. Therefore, the sequence  $\{\mathcal{N}_t(j)\}_{t=0}^{t_j^*}$  yields a filtration of the node set  $\mathcal{V}$  that satisfies Eq. (5). It follows that the set of filtrations of the form  $\{\mathcal{N}_t(j)\}_{t=0}^{t_j^*}$  (for  $j \in \mathcal{V}$ ) is a complete set of filtrations. We now prove consistency. Consider two realizations of a WTM contagion on a single network. Let  $\mathcal{N}_{t_i}(i) \subset \mathcal{V}$  denote the set of nodes that are adopters at time  $t_i$  for the  $i$ th realization, and let  $\mathcal{N}_{t_j}(j) \subset \mathcal{V}$  denote the set of nodes that are adopters at time  $t_j$  for the  $j$ th realization. To have consistency, it must be true that  $\mathcal{N}_{t_i+1}(i) \subseteq \mathcal{N}_{t_j+1}(j)$  if  $\mathcal{N}_{t_i}(i) \subseteq \mathcal{N}_{t_j}(j)$ . Suppose that  $t_i$  and  $t_j$  are times such that  $\mathcal{N}_{t_i}(i) \subseteq \mathcal{N}_{t_j}(j)$ , and consider the spreading that occurs for a WTM contagion during one time step. By definition, the update rule for each node is identical across all realizations of a WTM contagion. (In other words, for a node  $k$ , the fraction of infected neighbors  $f_k$  must surpass  $T$  for

adoption.) Additionally, for any node  $k \in \mathcal{V}$ , increasing the infection size can only increase  $f_k$ . Hence, if  $f_k > T$  for some node  $k$  when nodes  $\mathcal{N}_{t_i}(i)$  are infected, then  $f_k > T$  is also true if we instead consider a superset of  $\mathcal{N}_{t_i}(i)$  to be infected. Thus, the set  $\mathcal{N}_{t_{i+1}}(\hat{i})$  of adopters at time step  $t = t_i + 1$  must satisfy  $\mathcal{N}_{t_{i+1}}(\hat{i}) \subseteq \mathcal{N}_{t_{j+1}}(\hat{j})$ . The  $N$  realizations of a WTM contagion with seeds  $\mathcal{S}^{(\hat{j})} = \{\hat{j}\}$  (where  $\hat{j} \in \mathcal{V}$ ) thus produce to a complete and consistent set of filtrations on  $\mathcal{V}$  for which  $t_j^{(i)} = x_j^{(i)}$ . It follows that the activation times define a metric on the node set  $\mathcal{V}$ .  $\square$

## Supplementary Note 5 Algorithm for Constructing WTM Maps

In this section, we describe our algorithm for constructing WTM maps and discuss its computational complexity. We also conduct numerical simulations to illustrate the scaling of computation time with respect to network size  $N$  and mean node degree  $d = N^{-1} \sum_{i,j} A_{ij}$ . We thereby confirm that the typically observed computational cost of our algorithm scales quadratically with  $N$  and linearly with  $d$ . That is, for  $N$  nodes and  $M = Nd/2$  edges, the typical computational cost is  $\mathcal{O}(NM)$ .

### Algorithm and Computational Complexity

We begin by describing our algorithm for constructing a WTM map. (See the pseudocode in Algorithm 1 for a summary.) For a WTM map of a network with  $N$  nodes, we implement  $N$  realizations of a WTM contagion. We simulate the  $j$ th realization with cluster seeding by initializing the nodes in the contagion seed  $\mathcal{S}^{(j)} = \{j\} \cup \{k | A_{jk} \neq 0\}$  (i.e., node  $j$  and its neighbors) as infected and all other nodes as uninfected. The activation time for the seed nodes is  $t = 0$ . That is,  $x_j^{(i)} = 0$  for  $i \in \mathcal{S}^{(j)}$ . After initialization, we simulate a WTM contagion for time steps  $t = 1, 2, \dots$  until the dynamics reaches equilibrium. In other words, we reach a time step in which no new node becomes infected; this is guaranteed to occur at some time  $t < N$ . When considering the set of nodes that can become infected during a given time step  $t$ , it is sufficient to check only the subset of nodes  $i \in \mathcal{V}$  that are (1) not yet infected and (2) adjacent to a node that was infected during the previous time step (i.e., at time  $t - 1$ ). Therefore, as the contagion spreads, it is important to record which nodes adopt the contagion during each time step. Given such a list, upon reaching time step  $t$ , we examine the neighbors of all nodes that became infected at time  $t - 1$ . Any uninfected node  $i$  (among those neighbors) whose ratio  $f_i$  of infected neighbors to total neighbors satisfies  $f_i > T$  then becomes infected at time  $t$  (which we record as its activation time).

We now comment on the computational complexity of Supplementary Algorithm 1. There are  $N$  different contagions (because of cluster seeding centered at node  $j \in \mathcal{V}$ ). For each one, we need to calculate the activation time of every node  $i \in \mathcal{V}$ . Therefore, the computational complexity of computing a WTM map is at least  $\mathcal{O}(N^2)$ . Because we examine the neighbors of recently infected nodes at each time step, our algorithm also scales linearly with the node degree  $d = d^{(G)} + d^{(NG)}$  (which is identical for every node  $i \in \mathcal{V}$  in the experiments below), giving a total complexity of  $\mathcal{O}(N^2d)$ . However, there is scope to speed up the construction of WTM maps. If one constructs the dissimilarity matrix that encodes shortest paths between nodes (e.g., as required by Isomap [23]) using a “naive” method, then its computational complexity is also  $\mathcal{O}(N^2)$ . However, one can speed up the problem of computing shortest paths using Dijkstra’s algorithm [19], and we expect that similar improvements are possible for WTM maps.

Before we numerically support the  $\mathcal{O}(N^2d)$  computational complexity, we comment on the worst-case scenario, which has a complexity of  $\mathcal{O}(N^3)$ . This situation corresponds to a network in which every node is connected to every other node and exactly one node adopts the contagion at each time step for every contagion. Although such a scenario is technically feasible with general WTM contagion dynamics [20], this cannot arise in the WTM contagions that we study (and we also note that one can also analyze such a pathological situation using mean-field theory [21]) because we set  $T_i = T$  for all  $i \in \mathcal{V}$  in our experiments. Finally, although our implementation of Supplementary Algorithm 1 is sufficiently fast for the purposes of the present paper, we note that one can speed it up further by parallelizing it because the different initial conditions are independent of each other.

---

**Supplementary Algorithm 1** Construction of a WTM map with threshold  $T$  for a network with  $N$  nodes

---

```
1: for each node  $j \in \mathcal{V} = \{1, \dots, N\}$  do
2:   Initialize cluster seeding by infecting  $j$  and its neighbors; record their activation times as 0
3:   Run WTM contagion dynamics:
4:   while dynamics has not reached equilibrium do
5:     for  $i$  is a neighbor of a node that was infected during the last time step do
6:       if  $i$  is still uninfected then
7:         if fraction of activated neighbors  $f_i > T$  then
8:           infect node  $i$  and record its activation time
9:         end if
10:      end if
11:    end for
12:  end while
13: end for
```

---

## Numerical Investigation of Computational Cost

We implement Supplementary Algorithm 1 in both MATLAB and C++. In our discussion, we focus on our C++ implementation (which we have made publicly available, as we discussed in Sec. III A of the main text). We conduct numerical experiments to study the scaling behavior of the computational cost with respect to  $N$  and  $\langle d \rangle$ . All of the results that we report in the present section are mean values that we compute using 10 realizations for a particular choice of parameter values. We run these simulations on a computer with the following specifications: Debian GNU/Linux 7 operating system; 32 GB RAM memory; and 8 processor cores (Intel Core i7-4770 CPU @ 3.40GHz).

In Supplementary Fig. 10(a), we show the run times  $\delta t$  (in seconds) of our algorithm. These give computational costs for constructing WTM maps for noisy ring lattices with various sizes  $N \in [32, 31623]$ , which we construct while keeping the node degrees fixed at  $(d^{(G)}, d^{(NG)}) = (10, 2)$ . We show results for thresholds  $T = \{0.05, 0.2, 0.35, 0.45\}$ . Note for these values of  $T$  that the dependence of  $\delta t$  on  $N$  is much stronger than the dependence of  $\delta t$  on  $T$ . The symbols in Supplementary Fig. 10(a) give the observed computation times, and the solid lines give the inferred scaling behavior, which we assume takes the form  $\delta t = 10^\Gamma N^\zeta$  for some constants  $\zeta$  and  $\Gamma$ . In Supplementary Table 1, we summarize the fitted values for exponent  $\zeta$  and the prefactor  $\Gamma$  for various values of  $T$ . (As we discuss in the table caption, we use a least-squares fit.) We find that  $\zeta \approx 2$  for all  $T$ , supporting our claim of quadratic scaling behavior. We neglect the observed values of  $\delta t$  for  $N = 32$  for our fitting procedure because we are interested in the scaling as  $N \rightarrow \infty$ .

In Supplementary Fig. 10(b), we investigate the dependence of the computational cost on node degree  $d$ . In this experiment, we fix  $N = 2000$  and  $T = 0.35$  [which yields the largest values of  $\delta t$  in Supplementary Fig. 10(a)], and we vary  $d \in \{12, 24, \dots, 96\}$ . We plot the observed values of  $\delta t$  versus  $d$  for several choices of the ratio  $\alpha = d^{(NG)}/d^{(NG)}$ . For all values of  $\alpha$ , we observe positive scaling with  $d$  that we expect to be linear. As expected, we find that  $\delta t$  is much smaller when  $\alpha \geq 1/2$  than when  $\alpha < 1/2$ . For large values of  $\alpha$ , WTM contagions tend to either not spread at all (e.g., when  $T$  is large) or spread very quickly due to frequent ANC (e.g., when  $T$  is small). Therefore, the number of time steps that are necessary to reach equilibrium is small. This, in turn, yields a small value of  $\delta t$ .

In Supplementary Fig. 10(c), we further explore the dependence of  $\delta t$  on WFP and ANC by plotting

$\delta t$  versus threshold  $T$  for a noisy ring lattice with  $N = 10000$  nodes and  $(d^{(G)}, d^{(NG)}) = (2, 10)$ . (In this case,  $\alpha = 0.2$ .) First, note that there is a very large drop in  $\delta t$  near  $T = T_0^{(\text{WFP})} \approx 0.4167$  that corresponds to the bifurcation that separates the region in which the contagion spreads from the one in which it does not. For  $T \gtrsim 0.4167$ , the contagion spreads to just a few additional nodes (or no additional nodes), so it requires very few time steps for to reach an equilibrium state. For  $T \leq 0.4$ , the contagion spreads faster as  $T$  increases, which leads in turn to larger values of  $\delta t$ . Finally, note that there are several sharp jumps in  $\delta t$ ; these correspond to the bifurcations in the contagion dynamics [see Eqs. (7) and (13)].

## Supplementary Note 6 Additional Theory for Noisy Ring Lattices

In this section, we extend the bifurcation analysis that we presented in Sec. III B of the main text. In particular, we provide further details on our analysis for two contagion phenomena: *wavefront propagation* (WFP) along a network's underlying manifold and the *appearance of new clusters* (ANC) of a contagion due to transmission across non-geometric edges.

### Appearance of New Contagion Clusters (ANC)

ANC describes a contagion transmission in which a node becomes infected exclusively due to exposure via non-geometric edges. That is, the node's neighbors from geometric edges must not already be infected. As we discussed in Secs. I D and III B of the main text, we are able to describe this phenomenon with a sequence of critical thresholds:

$$T_k^{(\text{ANC})} \triangleq \frac{d^{(\text{NG})} - k}{d^{(\text{G})} + d^{(\text{NG})}}, \quad k = 0, 1, \dots, d^{(\text{NG})}, \quad (7)$$

where  $d^{(\text{G})}$  and  $d^{(\text{NG})}$ , respectively, denote a node's geometric and non-geometric degree for the noisy ring lattice with  $N$  nodes. (A node's "geometric degree" is its number of geometric stubs, that is, the number of its stubs that obey the original geometric space constraints, and a node's "non-geometric degree" is its number of non-geometric stubs.) For  $T \in [T_{k+1}^{(\text{ANC})}, T_k^{(\text{ANC})})$ , a node adopts a contagion if at least  $(d^{(\text{NG})} - k)$  non-geometric neighbors are infected. For  $T \geq T_0^{(\text{ANC})}$ , the contagion cannot spread exclusively by exposure to the contagion via non-geometric edges. In this section, we show by considering spreading exclusively on the subgraph that includes all nodes but only non-geometric edges that the rate of ANC of a WTM contagion increases as the contagion threshold  $T$  decreases.

We first consider the probability that a given node has exactly  $k$  infected non-geometric neighbors, given that  $q(t)$  of the  $N$  nodes are infected at time step  $t$ . First, consider the case  $k = 1$ , in which a node  $i$  has exactly one infected non-geometric neighbor. Given that node  $i$  has  $d^{(\text{NG})}$  non-geometric edges (which we label as  $e_1, \dots, e_{d^{(\text{NG})}}$ ), there are  $d^{(\text{NG})}$  possible outcomes with  $k = 1$ . For example,  $e_1$  is incident to an infected node and the remaining edges are incident to uninfected nodes,  $e_2$  is incident to an infected node and the remaining edges are incident to uninfected nodes, and so on. Recalling that we place non-geometric edges uniformly at random for the noisy ring lattice, the probability that edge  $e_1$  is incident to an infected node is  $\frac{q(t)}{N-1}$ , as there are  $q(t)$  such potential infected nodes and there are  $N - 1$  other nodes (because there are no self-edges). If edge  $e_1$  is incident to an infected node, then the probability that edge  $e_2$  is incident to an uninfected node is  $\frac{N-1-q(t)}{N-2}$ . If edges  $e_1$  and  $e_2$  are incident, respectively, to an infected node and an uninfected node, then the probability that edge  $e_3$  is incident to an uninfected node is  $\frac{N-1-q(t)-1}{N-3}$ . We can continue arguing similarly for the other edges. Taking into account that there are  $d^{(\text{NG})}$  possible outcomes in which the  $d^{(\text{NG})}$  edges are incident to exactly one infected node, the probability that a node has exactly one infected non-geometric neighbor is

$$P(1) = d^{(\text{NG})} \times \frac{q(t) \left( \prod_{k'=0}^{d^{(\text{NG})}-2} (N-1-q(t)-k') \right)}{\prod_{k'=0}^{d^{(\text{NG})}-1} (N-1-k')}. \quad (8)$$

More generally, the probability that a node has exactly  $k$  infected non-geometric neighbors is

$$P(k) = \binom{d^{(\text{NG})}}{k} \frac{\left(\prod_{k'=0}^{k-1} q(t) - k'\right) \left(\prod_{k'=0}^{d^{(\text{NG})}-1-k} N - 1 - q(t) - k'\right)}{\prod_{k'=0}^{d^{(\text{NG})}-1} (N - 1 - k')}. \quad (9)$$

For fixed  $d^{(\text{NG})} \ll N$  and  $q(t) = \mathcal{O}(N)$ , Eq. (9) simplifies to

$$P(k) \approx \binom{d^{(\text{NG})}}{k} \left[\frac{q(t)}{N}\right]^k \left[1 - \frac{q(t)}{N}\right]^{d^{(\text{NG})}-k}. \quad (10)$$

We now estimate the expected contagion size  $g(t)$  of a WTM contagion that spreads exclusively via ANC. In other words, we neglect exposures to the contagion from geometric edges, as we are assuming that they do not contribute to spreading. We define

$$k^{(\text{ANC})}(T) = \max \left\{ k' \mid T_{k'}^{(\text{ANC})} > T \right\}. \quad (11)$$

It follows that the minimum number of non-geometric neighbors that need to be infected for a node  $i$  to adopt the WTM contagion is  $(d^{(\text{NG})} - k^{(\text{ANC})})$ . Using Eq. (9) and Eq. (11), we estimate that the expected contagion growth satisfies

$$g(t+1) = g(t) + [N - g(t)] \sum_{k'=0}^{k^{(\text{ANC})}} P(d^{(\text{NG})} - k'), \quad (12)$$

where we calculate the expectation for  $g(t)$  over the ensemble of noisy ring lattices. We again stress that Eq. (12) estimates the size of a WTM contagion for ANC independent of WFP and does not account for the joint effect of spreading via both geometric and non-geometric edges. It therefore gives a lower bound for the size of the contagion [i.e.,  $q(t)$ ] for the regime that exhibits ANC but no WFP.

## Wavefront Propagation (WFP)

WFP describes the situation in which a contagion cluster expands because a node in its “boundary,” which we define as the set of nodes that are adjacent via a geometric edge to an infected node in the contagion cluster, becomes infected at time step  $t$ . In the main text, we found that WFP has the following sequence of critical thresholds:

$$T_k^{(\text{WFP})} \triangleq \frac{d^{(\text{G})}/2 - k}{d^{(\text{G})} + d^{(\text{NG})}}, \quad k = 0, 1, \dots, \frac{d^{(\text{G})}}{2}. \quad (13)$$

Assuming that the non-geometric edges of nodes in the contagion cluster’s boundary are incident to nodes that are not infected, a wavefront propagates with a speed of  $k+1$  nodes per time step for  $T \in \left[ T_{k+1}^{(\text{WFP})}, T_k^{(\text{WFP})} \right)$ .

For  $T \geq T_0^{(\text{WFP})}$ , there is no WFP. For a contagion that consists of a single cluster that is expanding via WFP in both directions along a noisy ring lattice, the size  $q(t)$  of the contagion (i.e., the number of nodes that have adopted the contagion) for time  $t \in \{0, 1, 2, \dots\}$  has a lower bound of

$$h(t) = \underbrace{(1 + d^{(\text{G})} + d^{(\text{NG})})}_{\text{seed nodes}} + 2k^{(\text{WFP})}t, \quad (14)$$

where

$$k^{(\text{WFP})} \triangleq 1 + \max\{k' \mid T_{k'}^{(\text{WFP})} > T\} \quad (15)$$

and the factor of 2 accounts for WFP in both directions along the ring.

Note that  $h(t)$  is a lower bound for  $q(t)$  because we have assumed that the non-geometric edges of nodes in the contagion cluster's boundary are incident to nodes that are not infected. This assumption is not always valid, so Eq. (14) is a lower bound because the invalidation of this assumption can only increase the rate of WFP. That is, nodes in a contagion cluster's boundary will adopt a contagion even if the number of geometric neighbors that are infected is smaller than what is required by Eq. (13).

Above we showed that the expected probability that a non-geometric edge of a node is incident to an infected node is  $q(t)/(N-1) \approx q(t)/N$ . Similarly, for a node with a non-geometric degree of  $d^{(\text{NG})}$ , the expected probability that none of its non-geometric edges are incident to an infected node is approximately  $[1 - q(t)/N]^{d^{(\text{NG})}}$ . This is therefore the probability that our WFP analysis given by Eq. (13) is valid for a given node at a given time step  $t$ . For large networks (i.e.,  $N \gg 1$ ) and early stages of a contagion (i.e.,  $q(t) \ll N$ ), the probability that our assumption is valid is approximately equal to 1. In this situation, Eqs. (13)–(15) accurately describe WFP (and the spread of the contagion). However, when  $q(t) \approx N$ , our assumption is almost certainly invalid, and we observe accelerated speeds of WFP. Interestingly, for large networks (i.e.,  $N \gg 1$ ), we find that such acceleration occurs infrequently early in a WTM contagion and that it occurs rather frequently towards the end of a contagion [i.e., just before  $q(t) \rightarrow N$ , which is when a contagion saturates a network]. Accelerated WFP is improbable [because  $q(t)$  is small, but  $N$  is large] in the early stages of a contagion on a large network. When  $t = 0$ , for example,  $q(0) = d^{(\text{G})} + d^{(\text{NG})} \ll N$ . However, during the late stages of a contagion [i.e.,  $q(t) \approx N$ ], accelerated WFP is very likely at every time step. Therefore, for small  $q(t)$ , Eq. (14) is both a lower bound for  $q(t)$  and an approximation for it. In general, the speed of WFP increases with time until it reaches an upper bound of  $d^{(\text{G})}/2$  nodes per time step. This bound corresponds to the situation in which all nodes that are incident via geometric edges to one side of a contagion cluster become infected during each time step. Note that there is no acceleration of WFP when  $k^{(\text{WFP})} = d^{(\text{G})}/2$ , as the wavefront is already propagating at its fastest rate.



## Supplementary Note 7 Extended Discussion of Point-Cloud Analyses

In this section, we provide further details on our approach to analyzing the point clouds that result from WTM maps. In particular, we provide a detailed discussion of the following three items:

1. The *Pearson correlation coefficient*  $\rho$ , which we use to investigate a point cloud’s geometry.
2. The *embedding dimension*  $P$ , which we use to investigate a point cloud’s dimensionality.
3. The *difference*  $\Delta = l_1 - l_2$  in lifetimes of the two most persistent 1-cycles [i.e., one-dimensional (1D) holes], which we use to investigate a point cloud’s topology.

We restrict our discussion of the above items for a point cloud that results from a regular WTM map  $\mathcal{V}sto\{\mathbf{x}^{(i)}\}$ , but one can apply the same techniques to any point cloud, including one that results from a reflected WTM map  $\mathcal{V}sto\{\mathbf{y}^{(i)}\}$  or a symmetric WTM map  $\mathcal{V}sto\{\mathbf{z}^{(i)}\}$ . (See Sec. I C of the main text for further discussion of these maps.)

We find for certain WTM contagion parameters that the structure of the point cloud that results from a WTM maps can reveal manifold structure in the original network and that one can quantify such structure using the values of  $\rho$ ,  $P$ , and  $\Delta$ . Importantly, one can thus use our approach to study not only manifold structure in networks but also the WTM contagion dynamics itself (e.g., uncovering the extent to which WFP dominates ANC or vice versa).

### Analysis of Geometry

Studying the geometry of a point cloud such as  $\{\mathbf{x}^{(i)}\}_{i \in \mathcal{V}}$  that results from a WTM map can reveal the extent to which the geometry of a WTM contagion follows the underlying geometry of a network. We investigate the extent to which the distance between two nodes in a point cloud that results from a WTM map relates to the distance between those nodes in the original metric-space embedding of the noisy geometric network. Specifically, we restrict our attention to noisy geometric networks in which the nodes  $\mathcal{V}$  have intrinsic locations  $\{\mathbf{w}^{(i)}\}_{i \in \mathcal{V}} \in \mathcal{M}$  on a manifold  $\mathcal{M} \subset \mathbb{R}^p$ . That is, they lie in a  $p$ -dimensional ambient space  $\mathbb{R}^p$  that we equip with the Euclidean norm  $\|\mathbf{w}\|_2 = \sqrt{\sum_{k=1}^p w_k^2}$ . We require the dimension  $p$  to be equal to the point cloud’s “embedding dimension”. In other words, there is no subspace of dimension smaller than  $p$  that one can define by a hyperplane that captures the manifold. Using the Euclidean metric, the distance between nodes  $i$  and  $j$  in the ambient space is

$$m(i, j) = \sqrt{\sum_{k=1}^p (w_k^{(i)} - w_k^{(j)})^2}. \quad (16)$$

We also use the Euclidean norm for the point cloud  $\{\mathbf{x}^{(i)}\} \in \mathbb{R}^N$  that results from a WTM map. The distance between node  $i$  and node  $j$  in such a point cloud is thus given by

$$m^{(\text{WTM})}(i, j) = \sqrt{\sum_{k=1}^N (x_k^{(i)} - x_k^{(j)})^2}. \quad (17)$$

Given two sets of distances  $m$  and  $m^{(\text{WTM})}$ , we compute the Pearson correlation coefficient

$$\rho = \frac{\sum_{i=1}^N \sum_{j=i+1}^N [m(i, j) - \bar{m}(i, j)] [m^{(\text{WTM})}(i, j) - \overline{m^{(\text{WTM})}}(i, j)]}{\sqrt{\sum_{i=1}^N \sum_{j=i+1}^N [m(i, j) - \bar{m}(i, j)]^2} \sqrt{\sum_{i=1}^N \sum_{j=i+1}^N [m^{(\text{WTM})}(i, j) - \overline{m^{(\text{WTM})}}(i, j)]^2}} \quad (18)$$

between all non-identical, unordered pairs  $(i, j) \in \mathcal{V} \times \mathcal{V}$ . Because  $i \neq j$  for distinct nodes, there are  $N(N-1)/2$  such pairs.

Note that calculating Eq. (18) requires the activation time  $x_j^{(i)}$  to be finite for all nodes  $i$  and realizations  $j$  of a WTM contagion. Unfortunately, this is not the case whenever there is a node that never becomes activated. Indeed,  $x_j^{(i)} = \infty$  for all nodes other than the seed if the threshold  $T$  is too large. (For instance,  $T \geq \max\{T_0^{(\text{WFP})}, T_0^{(\text{ANC})}\}$  is too large for the example of the noisy ring lattice.) For practical purposes, we set  $x_j^{(i)} = 2N$  in such cases, where we note that  $x_j^{(i)} \leq N-1$  for any numerical simulation in which node  $i$  eventually becomes infected. In Supplementary Note 1, we discuss other methods for handling situations with activation times of infinity.

## Analysis of Dimensionality

We study the dimensionality of a point cloud that results from a WTM map by exploring its *embedding dimension*. For a manifold  $\mathcal{M} \subset \mathbb{R}^p$ , we define the embedding dimension  $P$  as the minimum hyperplane dimension over all hyperplanes that span the manifold  $\mathcal{M}$ . Because a point cloud typically contains noise, which can potentially increase the dimensionality above that of an underlying manifold, we estimate embedding dimension using residual variance [22, 23].

Given the set of points  $\{\mathbf{x}^{(i)}\}_{i \in \mathcal{V}} \in \mathbb{R}^N$ , we consider each  $i \in \mathcal{V}$  and let  $\{\hat{\mathbf{x}}^{(i)}(p)\}$  denote the linear projection onto  $\mathbb{R}^p$  that we obtain from principal component analysis (PCA) [22, 23]. Let  $\rho^{(p)}$  denote the Pearson correlation coefficient that relates node-to-node distances  $m^{(\text{WTM})}$  from the original point cloud to node-to-node distances

$$m^{(p)}(i, j) = \sqrt{\sum_{k=1}^p [\hat{x}_k^{(i)}(p) - \hat{x}_k^{(j)}(p)]^2} \quad (19)$$

in the projected point cloud. It follows that  $\rho^{(p)}$  is given by Eq. (18) with the substitution  $m(i, j) \rightarrow m^{(p)}(i, j)$ .

The residual variance of such a linear dimension reduction is  $R_p = 1 - (\rho^{(p)})^2$ . We estimate the embedding dimension as the smallest dimension  $P$  such that the residual variance is (strictly) less than 0.05. That is,  $P = \min\{p | R_p < 0.05\}$ . For our calculations of embedding dimension, we only consider dimensions up to  $P = 20$ , as this simplifies the computational overhead of calculating  $P$ . Our motivation for this simplification (besides reducing computational cost) is that we are particularly interested in determining whether or not  $P$  is close to the known embedding dimension (e.g.,  $P = 2$  for the unit circle in  $\mathbb{R}^2$ , in which our noisy ring lattices are embedded).

## Analysis of Topology

In this section, we explain how to analyze the topology of a point cloud that results from a WTM map. We present our analysis for a general point cloud  $\mathcal{U} = \{\mathbf{u}^{(i)}\}_{i=1}^n \in \mathbb{R}^J$  (i.e., there are  $n$  points  $\mathbf{u}^{(i)}$  in  $J$  dimensions). We note for a typical WTM map, for which we map all  $N$  nodes based on  $N$  contagions, that one obtains a point cloud  $\{\mathbf{u}^{(i)}\}$  with  $n = J = N$ .

A set  $\mathcal{U}$  has a very simple topology. If  $\mathbf{u}^{(i)} \neq \mathbf{u}^{(j)}$  for  $i \neq j$ , then  $\mathcal{U}$  consists of  $N$  distinct connected components that correspond to the points  $\{\mathbf{u}^{(i)}\}$ . There are no 1-cycles in  $\mathcal{U}$ . To infer the topology of a meaningful underlying manifold (if present) that gives rise to a point cloud, we consider its topology across different spatial scales. In particular, we are interested in the topology of the sets

$$\mathcal{U}^{(r)} = \bigcup_{i \in \{1, \dots, n\}} \{\mathbf{u} \in \mathbb{R}^J : \|\mathbf{u}^{(i)} - \mathbf{u}\|_2 \leq r\} \quad (20)$$

for different values of  $r \in [0, \infty)$ . That is, we study the topology of sets that we construct as the union of radius- $r$  balls centered at points  $\mathbf{u}^{(i)} \in \mathcal{U}$ . Note that  $\mathcal{U}^{(0)} = \mathcal{U}$ . We choose to use the Euclidean norm, but it is also possible to use other norms.

We start with an example. In Supplementary Fig. 11, we show a noisy point cloud that we sample from a ring manifold. In particular, we sample the points uniformly from a unit circle in  $\mathbb{R}^2$ , and we add a small amount of noise to their locations in the embedding space  $\mathbb{R}^2$ . When  $r = 0$ , there are 10 distinct connected components, which correspond to the individual points. As we increase  $r$ , four of the components merge to create a 1-cycle [see Supplementary Fig. 11(b)]. As we continue to increase  $r$ , this 1-cycle fills in very soon after its birth. After it is filled in, another 1-cycle appears when  $r = 0.5$  [see Supplementary Fig. 11(c)]. This 1-cycle persists for a larger range of  $r$  values than the first 1-cycle, and it appears to correspond to a ring manifold that underlies the point cloud. This illustrates that one can study the topology of a point cloud by examining 1-cycles that persist across different spatial scales. To make this statement more quantitative, we employ tools from persistent homology [15–17].

For every set  $\mathcal{U}^{(r)}$ , one can assign homology groups  $H_c(\mathcal{U}^{(r)})$ , where  $c \in \{0, 1, 2, \dots\}$ . The rank  $\beta_c$  of the group  $H_c(\mathcal{U}^{(r)})$  counts the number of  $c$ -dimensional topological features that are present in  $\mathcal{U}^{(r)}$ . In particular,  $\beta_0$  counts the number of connected components,  $\beta_1$  counts the number of 1-cycles (which one can construe as a 1D hole or loop), and  $\beta_2$  counts the number of cavities [i.e., two-dimensional (2D) holes]. The fact that  $\mathcal{U}^{(r)} \subseteq \mathcal{U}^{(r')}$  for  $r \leq r'$  is very important. As we discussed earlier in this section, a sequence of sets with this property is a filtration. Thus, for any sequence  $\{r_i\}$  that satisfies  $r_i \leq r_{i+1}$  for  $i \in \{1, 2, \dots\}$ , the sequence of sets  $\{\mathcal{U}^{(r_i)}\}$  forms a filtration of  $\mathbb{R}^2$ . Examining changes of the topological features across the different elements of  $\{\mathcal{U}^{(r_i)}\}$  reveals multiple-scale topological features of the point cloud  $\{\mathbf{u}^{(i)}\}$ .

In the present paper, we are interested in understanding the birth and death of 1-cycles of  $\mathcal{U}^{(r)}$  as we vary  $r$ . The quantity  $\beta_1$  encodes such information, which one can summarize by drawing a *persistence diagram*. In Supplementary Fig. 12, we show the  $\beta_1$  persistence diagram for the point cloud in Supplementary Fig. 11. The diagram contains two points, which correspond to the two 1-cycles that we discussed previously. The horizontal (“birth”) axis of the point is the value of  $r$  at which the 1-cycle corresponding to this point first appears in  $\mathcal{U}^{(r)}$ , and the vertical (“death”) axis indicates when the 1-cycle is filled in. Enumerating the points  $i = 1, 2, \dots$  for every point  $i$  with coordinates  $(r_b(i), r_d(i))$  in the persistence diagram (where  $r_b$  denotes when a feature is born and  $r_d$  denotes when a feature dies), we define the “lifetime”  $l_i = r_d(i) - r_b(i)$ , and we denote the set of lifetimes of all points by  $L = \{l_1, l_2, \dots\}$  (which we order such that  $l_1 \geq l_2 \geq \dots$ ). Topological features with longer lifetimes (i.e., ones that are more persistent) indicate more dominant

features in a point cloud. In our example, there is one point with a very short lifetime that corresponds to a 1-cycle that arises for a single spatial scale due to the noisy sampling. The other point has a much larger lifetime, which indicates that its associated 1-cycle persists across many spatial scales. We thereby identify the ring structure of the sampled manifold. For the purpose of identifying whether or not a point cloud lies on a ring manifold, we summarize persistence diagrams by using the difference  $\Delta = l_1 - l_2$  between the most persistent lifetimes. Large values of  $\Delta$  correspond to persistence diagrams that consist of a single point with a large lifetime, as we expect for a point cloud that lies on a ring manifold.

In practice, computing the persistent homology of a set  $\mathcal{U}^{(r)}$  is complicated. However, the so-called ‘‘Nerve Theorem’’ [24] guarantees that the homology of  $\mathcal{U}^{(r)}$  is the same as the homology of a corresponding Čech complex, which simplifies analysis but is computationally expensive to construct. Therefore, we study an approximation of the Čech complex that is known as the Vietoris-Rips complex. For a given point cloud  $\mathcal{U} = \{\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots, \mathbf{u}^{(n)}\} \in \mathbb{R}^J$  and  $r \in \mathbb{R}$ , the *Vietoris-Rips complex*  $\text{VR}^{(r)}$  consists of the simplices  $(\mathbf{u}^{(s_1)}, \mathbf{u}^{(s_2)}, \dots, \mathbf{u}^{(s_k)})$  such that  $\|\mathbf{u}^{(s_i)} - \mathbf{u}^{(s_j)}\|_2 \leq r$  for all  $s_i$  and  $s_j$ . In the present paper, we are interested only in identifying the 1-cycles in  $\text{VR}^{(r)}$ , so it is sufficient for us to use only 0-simplices (i.e., points), 1-simplices (i.e., line segments), and 2-simplices (i.e., triangles).

To compute persistent homology, we use the software package PERSEUS [25] (version 3.0 Beta), and we also check some of our results using the JAVAPLEX Persistent Homology Library [26]. To construct Vietoris-Rips filtrations  $\text{VR}^{(r)}$  for a point cloud that results from a WTM map (e.g.,  $\{\mathbf{x}^{(i)}\}$ ), we use Eq. (17) to define distances between points. As an input to PERSEUS, we use the dissimilarity matrix in which the entry in the  $i$ th row and  $j$ th column encodes the distance between nodes  $i$  and  $j$  given by Eq. (17).

In Supplementary Fig. 13, we study  $\beta_1$  persistence diagrams for point clouds that result from the application of WTM maps to noisy ring lattices. We thereby reveal the absence versus presence of 1-cycles in the point cloud. We analyze the  $\beta_1$  persistence diagrams for several values of the WTM threshold  $T \in [0, 0.5]$  and several choices for non-geometric degrees  $d^{(\text{NG})} \in [0, 20]$  for networks with  $N = 200$  nodes and a geometric degree of  $d^{(\text{G})} = 20$  (which implies that  $\alpha = d^{(\text{NG})}/d^{(\text{G})} \in [0, 1]$ ). A red diamond in Supplementary Fig. 13 represents the point that corresponds to the most persistent 1-cycle. We indicate the second-most persistent 1-cycle using a yellow square, and we mark the remaining points in the persistence diagram using white circles. If there is only one dominant 1-cycle, then the separation between the red diamond and the other points is large. To measure this separation, we calculate  $\Delta = l_1 - l_2$ , where  $l_1$  and  $l_2$  are the lifetimes of the dominant and the second most dominant 1-cycle, respectively. The background coloration reflects the value of  $\Delta$ . To construct a filtration using various values of  $r$ , we consider 100 evenly-spaced values of  $r$  that range from 0 up to the maximum distance  $r_{\max} \triangleq \max_{i,j \in \mathcal{V}} \|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\|_2$  between any two points. For our plots, we normalize all  $r$  values by  $r_{\max}$ , so  $\Delta \in [0, 1]$ . It follows that  $\Delta \approx 1$  indicates the presence of the ring topology, whereas small values of  $\Delta$  indicates its absence.

## Supplementary Note 8 Complex Contagions on a Ring Manifold

In this section, we give results for numerical experiments in which we study the geometry, dimensionality, and topology of point clouds that result from the application of symmetric WTM maps  $\mathcal{V}_{sto}\{\mathbf{z}^{(i)}\}$  to noisy geometric networks generated by network families **(a)**–**(d)**, which we defined in Supplementary Note 3. We thereby reveal the extent to which a WTM contagion exhibits WFP that follows the underlying ring manifold (i.e., the extent to which spreading occurs across a network subgraph that contains exclusively geometric edges) versus ANC. In particular, WFP is more prevalent than ANC when one can identify the properties of the underlying manifold in the point cloud that results from a WTM map.

To give some perspective for our numerical experiments, we compare our results for point clouds produced by WTM maps to results from two well-known methods of mapping network nodes as a point cloud: a Laplacian eigenmap [27] and Isomap [23]. In particular, we consider a 2D Laplacian eigenmap in which we map each node  $i$  to  $[v_i^{(2)}, v_i^{(3)}]^T \in \mathbb{R}^2$ , where  $\mathbf{v}^{(j)}$  is the eigenvector that corresponds to the  $j$ th eigenvalue  $\lambda_j$  of the unnormalized Laplacian matrix  $L$  (i.e.,  $L\mathbf{v}^{(j)} = \lambda_j\mathbf{v}^{(j)}$ ) and we have ordered the eigenvalues so that  $0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_N$ . The unnormalized Laplacian matrix has the form  $L = \text{diag}(d_1, d_2, \dots, d_N) - A$ , where  $d_i = d_i^{(G)} + d_i^{(NG)}$  is the total degree of node  $i$  and  $A$  is the adjacency matrix. As we discussed in Sec. I C of the main text, Isomap entails mapping network nodes based on the shortest paths between nodes. It corresponds to a WTM map with  $T = 0$  if we initialize the contagions with node seeding rather than cluster seeding. As we will see, when assessing the extent to which point clouds that result from WTM maps resemble the underlying ring manifold, we typically find a range of threshold values for which the geometry, dimensionality, and topology of the manifold is more apparent in WTM maps than for Laplacian eigenmap and Isomap methods. For other threshold values, the manifold is less apparent for WTM maps than for the other methods.

Note that Laplacian eigenmaps and Isomap were introduced originally for the purpose of nonlinear dimension reduction of point-cloud data rather than for network analysis. They were developed to map a high-dimensional point cloud to a network and then to map that network to a low-dimensional point cloud. Therefore, applying a Laplacian eigenmap or Isomap directly to a network—especially one that is unweighted—is different from what they were designed to do. In particular, for networks that arise from high-dimensional data—e.g., ones with nodes that are connected to each other by applying a  $k$ -nearest-neighbor algorithm—one often weights network edges based on distances in the original, high-dimensional point cloud. Incorporating such additional information can, of course, improve the results of dimension reduction (e.g., when attempting to “learn” manifold attributes such as topology, geometry, and dimensionality). Finally, when considering dimension reduction such as manifold learning in networks (i.e., rather than point clouds), one should determine the approach to dimension reduction (e.g., whether the algorithm is based on diffusion, shortest paths, or contagion dynamics) based on the application at hand. (For example, one might be more interested in conservative processes in some situations and in non-conservative processes in others.)

### Numerical Results for Geometry

In this section, we compare the geometry of symmetric WTM maps for networks in the families **(a)**–**(d)**, which we defined in Supplementary Note 3, via calculating a Pearson correlation coefficient  $\rho$  to compare WTM distances to distances in an underlying manifold. We also investigate the effects on WTM maps of varying the mean geometric and non-geometric degrees and the network size  $N$  (when we hold other parameters constant). We show our results in Supplementary Figs. 14–16. Panels **(a)**–**(d)**, respectively, give

our results for network families **(a)**–**(d)**. Unless we indicate otherwise, we show results for one network from each family in these and subsequent figures.

In Supplementary Fig. 14, we plot  $\rho$  for point clouds that result from symmetric WTM maps for the  $(T, \alpha)$  parameter plane. The solid and dashed curves yield approximate bifurcation curves, which we obtain from Eqs. (3) and (4) with  $\delta_i^{(G)} = \delta_i^{(NG)} = 0$ . Note that panel (a) depicts similar information as Supplementary Fig. 6(a) of the main text, although the results that we now show are for a larger network with larger degrees. For all panels, the curve given by  $T_0^{(WFP)}$  agrees very well with a relatively abrupt transition that one can observe by examining the geometry of the WTM maps via the coefficient  $\rho$ . In contrast, the curve for  $T_0^{(ANC)}$  appears to not be as closely related to  $\rho$ . To illustrate this, fix  $\alpha = 0.25$  and consider increasing values of  $T$  in any panel. As  $T$  surpasses  $T_0^{(WFP)}$ , there is a large drop in the value of  $\rho$ . By contrast,  $\rho$  changes only slightly when we cross  $T \approx T_0^{(ANC)}$ . Comparing the four panels to each other, we find that the agreement between  $T_0^{(WFP)}$  and the observed shifts in  $\rho$  decreases as the node degrees become more heterogeneous. In particular, the transition that is imposed by  $T_0^{(WFP)}$  appears to shift to smaller values of  $T$ , so the heterogeneities that we introduce in network families **(b)**–**(d)** mostly affect the regime in which  $T \approx T_0^{(WFP)}$ . However, the qualitative behavior of WTM contagions in the  $(T, \alpha)$  parameter plane is similar for all four families of networks.

In Supplementary Fig. 15, we study the effect on  $\rho$  for symmetric WTM maps when we increase the mean node degrees  $\langle d_i^{(G)} \rangle$  and  $\langle d_i^{(NG)} \rangle$ . Fixing  $\alpha = 1/3$ , we plot  $\rho$  as a function of the threshold  $T$ . This amounts to examining vertical cross sections of the four panels in Supplementary Fig. 14. We study the effect of varying mean node degree by showing results for  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (6, 2)$  (red triangles),  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (12, 4)$  (blue squares), and  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (24, 8)$  (magenta  $\times$  symbols). We also show results for a 2D Laplacian eigenmap [27] (horizontal dashed lines) and Isomap [23] (horizontal dotted lines) applied to our (unweighted) networks. Comparing the  $\rho$  values for the WTM maps with various mean degrees, we note that increasing the mean degrees smoothens the dependence of  $\rho$  versus  $T$ . Specifically, for smaller values of  $T$  (e.g., for  $T < 0.3$ ), the discontinuous jumps in  $\rho$  become smaller as the mean degrees increase. Interestingly, increasing heterogeneity in the node degrees also smoothens the curves of  $\rho$  versus  $T$ . For example, the curves are smoother for network families **(b)**–**(d)** than they are for family **(a)**. Additionally, note in all panels that we observe an abrupt drop in  $\rho$  for  $T \approx T_0^{(WFP)} = 1/(2 + 2\alpha) = 3/8$ . However, in more heterogeneous situations, this abrupt drop can shift to smaller values of  $T$ . This is most apparent when comparing the four panels for  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (6, 2)$  (red triangles): the drop occurs at  $T = 3/8 = 0.375$  in panel (a), whereas it occurs at approximately at  $T \approx 0.29$  in panels (b)–(d). To contrast this large shift, when comparing the panels for  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (24, 8)$  (magenta  $\times$  symbols), we observe that the change is smaller [i.e., the abrupt drop in  $\rho$  is at  $T = 0.375$  in panel (a), whereas it occurs at  $T \approx 0.35$  in panel (d)]. Finally, note in all panels that  $\rho$  increases as we increase the mean degrees, and we observe similar increases in  $\rho$  for the Laplacian-eigenmap and Isomap algorithms. Thus, in this series of experiments, increasing mean node degree improves the ability of the maps to translate the geometry of the underlying manifold to the resulting WTM point cloud.

In Supplementary Fig. 16, we study the geometry of symmetric WTM maps by plotting  $\rho$  versus  $T$  for networks of various sizes  $N$ . We fix  $\langle d_i^{(G)} \rangle = 24$  and  $\langle d_i^{(NG)} \rangle = 8$  (that is,  $\alpha = 1/3$ ) and plot  $\rho$  versus the threshold  $T$  for  $T \in [0, 0.6]$ . In each panel, we show results for several choices of network size  $N \in [200, 2000]$  to illustrate how  $\rho$  depends on  $N$ . As  $N$  increases, we observe that  $\rho$  systematically decreases for WTM maps that correspond to contagions in which WFP is not the dominant phenomenon. However, for WTM maps in which WFP dominates (e.g., when  $T \in [0.2, 0.25]$ ), we find that  $\rho$  remains

above 0.85. This provides strong evidence that, for this parameter regime, WTM maps translate the geometry of the underlying ring manifold to the resulting point cloud for a wide range of network sizes (with the other parameters held constant). One does *not* obtain such independence with network size when using a Laplacian eigenmap or Isomap. In those cases, we find that  $\rho$  systematically decreases as  $N$  increases (with the other parameters held constant).

## Numerical Results for Dimensionality

In this section, we examine the dimensionality of point clouds that result from symmetric WTM maps that we apply to networks on a ring manifold. As we discussed in Sec. III E and Supplementary Note 7, we study their “embedding dimension”  $P$ , which we define for a point cloud to be the smallest dimension  $p$  such that the residual variance  $R_p$  for the projection onto  $\mathbb{R}^p$  is small. In practice, we use PCA for such projections, and we specify “small” as being (strictly) less than 0.05. (In other words, we lose less than 5% of the variance after the projection.) Importantly, if the point cloud is a noisy sample of points on a manifold, then  $P$  is an approximation for the embedding dimension of the manifold.

We show our results for embedding dimension of point clouds resulting from WTM maps in Supplementary Figs. 17–18. Panels (a)–(d), respectively, give our results for network families (a)–(d). In Supplementary Fig. 17, we plot  $P$  in the  $(T, \alpha)$  parameter plane for networks with  $N = 200$  nodes, mean geometric degree  $\langle d_i^{(G)} \rangle = 20$ , and various mean non-geometric degrees  $\langle d_i^{(NG)} \rangle$ . We also plot the approximate bifurcation curves given by Eqs. (3) and (4) with  $\delta_i^{(G)} = \delta_i^{(NG)} = 0$ . Note that panel (a) is similar to the plot in Supplementary Fig. 6(b) of the main text. We observe in all panels that WTM maps for the contagion regime that we expect to exhibit WFP but no ANC yield point clouds  $\{\mathbf{z}^{(i)}\}$  with a small embedding dimension of  $P \approx 2$ . This result is expected, because a ring manifold is exactly the unit circle in  $\mathbb{R}^2$ . That is, it is a one-dimensional manifold that requires at least two dimensions to be embedded in a Euclidean space. Note that this low dimensionality persists into the regime that we expect to exhibit both WFP and ANC, although the embedding dimension  $P$  increases as one moves away from the regime exhibiting WFP and no ANC.

In Supplementary Fig. 18, we continue our investigation of the dimensionality of point clouds that result from the application of symmetric WTM maps to networks on a ring manifold by showing their embedding dimension  $P$  as a function of threshold  $T$ . One can construe the curves of  $P$  versus  $T$  as a vertical cross section of the contour plots in Supplementary Fig. 17; we show results for several choices of mean node degrees:  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (6, 2)$  (red triangles),  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (12, 4)$  (blue squares), and  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (18, 6)$  (magenta  $\times$  symbols). We also show values (horizontal dotted lines) of  $P$  versus  $T$  for the point clouds that we obtain by applying Isomap [23] to the networks. We obtain horizontal lines because Isomap does not include any dependence on  $T$ . We do not investigate the dimensionality of the 2D Laplacian eigenmaps, as we fix their dimension to 2 in our study.

Note that the curves in panels of Supplementary Fig. 18 are rather similar to each other. In particular, for all panels, we find the smallest embedding dimension  $P$  for the regime in which we expect a WTM contagion to exhibit WFP without ANC [i.e., for  $T \in (1/4, 3/8)$ ]. Additionally, we consistently identify the correct embedding dimension (i.e.,  $P = 2$ ) for this regime as long as mean degrees are sufficiently large (e.g., see the magenta  $\times$  symbols). For smaller mean degrees, we still observe that  $P$  is small for a similar range of the threshold  $T$ . However, the curves of  $P$  versus  $T$  tend to suggest that smaller mean degrees lead to larger embedding dimensions in our numerical experiments. For Isomap (in which we map nodes based on shortest paths), we observe in our experiments that the embedding dimension  $P$  is always at least 10. Additionally, the embedding dimension  $P$  for Isomap appears to decrease systematically as the

mean degrees increase. Thus, using shortest paths to map nodes for network families **(a)**–**(d)** leads to point clouds with a dimensionality that is higher than  $P = 2$ ; however, it might be possible to recover the correct embedding dimension of a ring manifold when the mean degrees are sufficiently large (keeping all other parameters fixed). Finally, note that  $P \leq 20$  in all panels. Recall that this is the maximum value of  $P$  that we can observe because it is the largest projection that we consider.

## Numerical Results for Topology

In this section, we study the topology of point clouds that result from symmetric WTM maps applied to noisy geometric networks on a ring manifold. As we discussed in Sec. III F and Supplementary Note 7, we examine the difference  $\Delta = l_1 - l_2$  between the largest lifetimes for 1D features (i.e., 1-cycles). We determine the persistence of these 1-cycles across spacial scales using a Vietoris-Rips filtration of the point cloud [15–17]. We normalize the difference in lifetimes so that  $\Delta \in [0, 1]$ . We show our results in Supplementary Figs. 19–20. Panels (a)–(d), respectively, give our results for network families **(a)**–**(d)**.

In Supplementary Fig. 19, we plot  $\Delta$  in the  $(T, \alpha)$  parameter plane. We show results for networks with  $N = 200$  nodes, mean geometric degree of  $\langle d_i^{(G)} \rangle = 20$  and various mean non-geometric degrees  $\langle d_i^{(NG)} \rangle \in [0, 20]$ . In each panel, the solid and dashed curves indicate, respectively, our approximate bifurcation curves from Eqs. (3) and (4) with  $\delta_i^{(G)} = \delta_i^{(NG)} = 0$ . Note that panel (a) is similar to Supplementary Fig. 6(c) from the main text. Variations in  $\Delta$  appear to correspond closely with the theoretical curves. For example, fixing  $\alpha < 0.5$  and increasing  $T$ , we observe an increase in  $\Delta$  as  $T$  surpasses  $T_0^{(ANC)}$  and a decrease in  $\Delta$  as  $T$  surpasses  $T_0^{(WFP)}$ . For all four network families,  $\Delta$  is largest for WTM maps that correspond to the contagion regime that we predict to be characterized by WFP without ANC. By comparing the panels, we see that the identifiability of the underlying ring topology (as indicated by  $\Delta \approx 1$ ) decreases as we increase the heterogeneity in the nodes' degrees. For example, panel (a) includes parameter values  $(T, \alpha)$  for which  $\Delta > 0.9$ , but  $\Delta < 0.6$  in panel (d) for the same portion of the parameter plane  $(T, \alpha)$ .

In Supplementary Fig. 20, we continue our investigation of the topology of the point clouds that result from symmetric WTM maps by fixing  $\alpha = 1/3$  and plotting  $\Delta$  as a function of  $T$ . One can construe these curves of  $\Delta$  versus  $T$  as examining vertical cross sections from the panels in Supplementary Fig. 19 with several choices of mean degrees:  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (6, 2)$  (red triangles),  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (12, 4)$  (blue squares), and  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (24, 8)$  (magenta  $\times$  symbols). As before, the networks have  $N = 200$  nodes. Observe in Supplementary Fig. 20 that  $\Delta$  tends to decrease as the heterogeneity of the network increases. For example, the values of  $\Delta$  in panels (b) and (c) tend to be smaller than those in panel (a), and the  $\Delta$  values in panel (d) tend to be even smaller than those in panels (b) and (c). This decrease makes it harder to successfully identify the ring topology in the point clouds. This is most evident for the curves that correspond to  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (6, 2)$  (red triangles). Although we observe large values of  $\Delta$  in panels (a) and (b) for the point clouds for the contagion regime that we predict to exhibit WFP without ANC [i.e.,  $T \in (1/4, 3/8)$ ], we find that the values of  $\Delta$  for this regime are much smaller in panels (c) and (d). In this experiment,  $\Delta$  does not depend on mean node degrees in a simple manner. In network family **(a)**, for example, when comparing the  $\Delta$  versus  $T$  curve for  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (12, 4)$  to that for  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (6, 2)$ , we observe larger  $\Delta$  values when increasing the mean degrees. However, restricting out attention to the range  $T \in (1/4, 3/8)$ , the curve of  $\Delta$  versus  $T$  for  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (24, 8)$  yields  $\Delta$  values that are smaller than those for  $(\langle d_i^{(G)} \rangle, \langle d_i^{(NG)} \rangle) = (12, 4)$ . This nontrivial behavior might be due to the relatively small differences in magnitude between the mean node degrees and the network size



( $N = 200$ ) that we use for this experiment. In this experiment, we also compute  $\Delta$  for Isomap, and we find that  $\Delta \approx 0$  in all cases. We thus omit these results from Supplementary Fig. 20.

## Non-Uniform Sampling of a Ring Manifold

In our numerical experiments thus far, we have investigated symmetric WTM maps for four families of noisy geometric networks on a ring manifold. (See Supplementary Note 3 for their descriptions.) Network families **(c)** and **(d)** allow heterogeneity in the node locations along a ring manifold through the placement of nodes via unevenly-spaced angles  $\{\theta_i\}$  along the unit circle. Recall that each node  $i$  has an associated angle  $\theta_i = \frac{2\pi i}{N} + \delta\theta_i$ , where we draw  $\delta\theta_i \sim \mathcal{N}(0, (s\frac{2\pi}{N})^2)$  from a Gaussian distribution with a standard deviation of  $s\frac{2\pi}{N}$ . Note that  $\frac{2\pi}{N}$  is the spacing between the  $N$  nodes if they are spaced uniformly on the ring. Consequently, by varying the parameter  $s$ , one can tune the level of heterogeneity in node location and thus the heterogeneity of the geometric degrees  $\{d_i^{(G)}\}$ . Recall that  $s \rightarrow \infty$  corresponds to sampling locations on the unit circle uniformly at random. In our previous experiments, we let  $s = 1/2$  for network families **(c)** and **(d)**. In this section, we investigate the effect of varying  $s$ . Because  $s > 0$  introduces heterogeneity in the geometric degrees, we consider both the case in which the nodes' non-geometric degrees are identical and the case in which they are heterogeneous. That is, the networks that we now consider are generalizations of network families **(c)** and **(d)**, but we now also vary the level of heterogeneity in the geographic spacing of nodes on the ring.

In Supplementary Fig. 21, we show results for the (left column) geometry, (center column) dimensionality, and (right column) topology of symmetric WTM maps, where we fix  $\alpha = 1/3$  and  $N = 200$  and we vary the threshold  $T$ . We consider networks with  $N = 200$  nodes, mean geometric degree of  $\langle d_i^{(G)} \rangle = 24$ , mean non-geometric degree of  $\langle d_i^{(NG)} \rangle = 8$ , and  $s \in \{0, 1/2, 1, 3/2, 2, 5/2, \infty\}$ . The top row corresponds to generating noisy geometric edges so that every node has exactly  $d_i^{(NG)} = \langle d_i^{(NG)} \rangle$  non-geometric edges, and the bottom row corresponds to generating noisy geometric edges uniformly at random so that the non-geometric degree  $d_i^{(NG)}$  of a node  $i$  is a binomially-distributed random variable. See the descriptions of the network families in Supplementary Note 3. Using horizontal dashed lines, we show results for the mapping of nodes for Isomap (i.e., based on shortest paths). We omit these results from panels (c) and (f), because we obtain  $\Delta \approx 0$  in these cases. The dashed lines in panels (a) and (d) give values of  $\rho$  for a 2D Laplacian eigenmap. (It is 2D by construction, so we do not investigate its embedding dimension  $P$ .)

Increasing network heterogeneity by increasing  $s$  has a significant effect on the structure of the point clouds that result from symmetric WTM maps. For example, we see in panels (a) and (d) that increasing  $s$  shifts the abrupt drop-off in the Pearson correlation coefficient  $\rho$ , which originally occurs near its expected value of  $T_0^{(WFP)} = 3/8$ , to progressively smaller values of  $T$ . In fact, we see in all panels that increasing  $s$  causes the curves of  $\rho$  versus  $T$  to shift to the left. Additionally, in panels (a) and (d), we see for sufficiently large  $s$  that there is a regime in which  $\rho$  is small for all threshold values  $T$ . In panels (b) and (e), we still obtain regimes in which the WTM maps are low-dimensional (i.e.,  $P \approx 2$ ). However, as  $s$  increases, the range of  $T$  values for which  $P$  indicates low dimensionality becomes smaller and shifts to the left. In panels (c) and (f), one can also observe that the ability to identify the ring topology becomes more difficult with increasing  $s$ . In panel (c), we obtain large 1-cycle lifetimes  $\Delta$  when  $T \in (1/4, 3/8)$  for  $s = 0$ ; this provides strong evidence that the point cloud lies on a ring manifold. For small  $s$  (e.g.,  $0 \leq s \leq 3/2$ ), we also obtain large values of  $\Delta$ , but the range of thresholds  $T$  that produce large  $\Delta$  are smaller and have shifted to the left. However, when  $s$  is large (e.g.,  $s = \infty$ ),  $\Delta$  remains small for all threshold values  $T$  in panels (c). There is even less evidence of the ring topology in panel (f), as  $\Delta$  remains small for all values of  $s$  and  $T$ .

## Summary of Experiments with a Ring Manifold

We have conducted an extensive investigation of the geometry, dimensionality, and topology of symmetric WTM maps for several families of noisy geometric networks on a ring manifold. We now briefly summarize our results.

We demonstrated that the structure (e.g., geometry, dimensionality, and topology) of WTM maps depends strongly on the contagion threshold  $T$  and the network parameters (e.g., the number of nodes  $N$ , the geometric degrees  $\{d_i^{(G)}\}$ , and the non-geometric degrees  $\{d_i^{(G)}\}$ ). Consequently, the extent to which a WTM contagion exhibits wavefront propagation (WFP) versus the appearance of new clusters (ANC) of contagions also depends on these parameters. Bifurcation analysis did a good job of predicting which parameter regimes have similar point-cloud structures. This is particularly evident in the  $(T, \alpha)$  parameter plane in Supplementary Figs. 14, 17, and 19, where we observed that the geometry, dimensionality, and topology of WTM maps align well with our theoretical predictions for the occurrence of bifurcations in the dynamics of a WTM contagion. We found such agreement even for networks with heterogeneity in geometric degrees, non-geometric degrees, and/or node locations along a ring manifold. As we discussed in Supplementary Note 3, we interpret our bifurcation analysis for the noisy ring lattice as an approximate bifurcation analysis for the networks with heterogeneous structures. As expected, we also observed that the accuracy of our approximation increases as the mean node degrees increase. However, its accuracy is sensitive to a variety of factors—including the threshold  $T$ , the network size  $N$ , and the particular type of structural heterogeneity in the network. In many of our numerical experiments, we compared the structure of point clouds that result from WTM maps to those that result from a 2D Laplacian eigenmap [27] and Isomap [23], which map the network nodes based on diffusion dynamics and shortest paths, respectively. Our approach provides a nice complement to these methods.

## Supplementary Discussion

In this Supplementary Discussion, we further consider the implications of our study for three research areas that have diverse motivations and goals but which share a common interest in understanding spreading processes on networks.

### High-Dimensional Data Analysis of Contagions and Other Dynamics

Research on network epidemiology [28–31] underscores the importance of the perspective that we have taken in the present paper. For example, Brockmann and Helbing [28] recently defined node-to-node distances based on a stochastic model for contagions that takes into account human mobility patterns in the worldwide airline network, and they reported that such a notion of distance did a good job of predicting global contagions. In their study, Brockmann and Helbing reported that node-to-node distances are insensitive to the contagion parameters in their model. By contrast, we find that the geometry, dimensionality, and topology of contagions depends sensitively on the contagion parameters (e.g., the threshold  $T$ ) of the WTM. This appears to arise from the thresholding process, so we expect it to be relevant for complex contagions in general because of the importance of social reinforcement [32–35].

Our perspective can be applied to study other spreading processes [21], where it has the potential to offer insights into phenomena such as information seeding [36] and targeted immunization [37, 38]. Moreover, a large variety of other processes—including some of the most heavily investigated dynamical processes (e.g.,  $k$ -core percolation and other types of percolation) [21, 39], more intricate complex-contagion models [40], and even some local methods for community detection [41]—also satisfy filtration conditions that are based on node states and the dynamics of such states. One can thus construct contagion maps for these processes and study them using the approach that we have illustrated. Computational homology offers a promising (and novel) approach for studying all of those situations.

### Dimension Reduction of Networks

In the present paper, we used the fact that WTM contagions satisfy a filtration condition. This makes it possible to study networks from the perspective of computational topology [42–45]. One can thus construct a metric space based on when nodes adopt a contagion for different choices of initial conditions. (See Supplementary Note 4.) WTM contagions thereby allow the simultaneous study of network topology, geometry, and dimensionality. Such manifold learning has numerous applications, including inference of missing and spurious edges [6–10], efficient routing of information [46, 47], and identification of attributes that are responsible for edge formation [48]. To provide a step in this direction, in Supplementary Note 2, we compared the denoising of networks via WTM maps—a “global” approach for identifying spurious edges—to a popular “local” approach based on the statistics of subgraphs [4].

An important future direction is to improve the computational efficiency of constructing contagions maps. As we discussed in Supplementary Note 5, the typical computational complexity for our construction of a WTM map with all possible initial conditions with clustering seeding is currently  $\mathcal{O}(NM)$ , where  $M$  is the number of edges in a network. Approximation schemes based on ideas such as network sampling [41] and random projections [49] offer promising approaches for improving computation speed.

## Dimension Reduction of Point-Cloud Data

Although we focused on manifold structure in networks, our approach extends naturally to point-cloud data (e.g., images, videos, and time series)—the traditional setting for manifold learning—if one first infers a proximity network using, for example, a  $k$ -nearest neighbor distance thresholding [27, 50–52]. In this endeavor, a central pursuit has been the development of techniques that are robust to noise [51–53]. It is well-known that diffusion distances are more robust than shortest-path distances to noisy edges, so maps that are based on diffusion [27, 50] can be preferable to the Isomap algorithm [23] for noisy data [53]. However, noisy edges can still be problematic for diffusion distances, so some techniques attempt to denoise a network prior to mapping it [7]. The robustness to noisy edges for WTM maps with contagions dominated by WFP makes them appealing, and it would be interesting to explore applications with noisy data.

An important distinction of WTM maps from prior work is that our research is based on nonlinear and nonconservative dynamics (in particular, on complex contagions) rather than on linear and conservative dynamics such as diffusion (e.g., random walks) [6, 7, 27, 50–52]. These different classes of dynamics can behave very differently, and it is known that they give very different answers for questions like which nodes are most important [54] and what network structures constitute bottlenecks to such dynamics [55] (which is closely related to which network structures yield dense communities of nodes [41]). Comparing WTM maps to Laplacian eigenmaps [27] and Isomaps [23] (see Supplementary Note 8) illustrates that these different dynamics lead to differences in the results of dimension reduction. It is thus important to explore dynamics other than diffusion for the analysis of point-cloud data.

## Supplementary References

- [1] Lee, S. H. & Holme, P. Exploring maps with greedy navigators. *Phys. Rev. Lett.* **108**, 128701 (2012).
- [2] Rombach, M. P., Porter, M. A., Fowler, J. H. & Mucha, P. J. Core-periphery structure in networks. *SIAM J. Appl. Math.* **74**, 167–190 (2014).
- [3] Gleeson, J. P. & Cahalane, D. J. Seed size strongly affects cascades on random networks. *Phys. Rev. E* **75**(5), 056103 (2007).
- [4] Goldberg, D. S. & Roth, F. P. Assessing experimentally derived interactions in a small world, *Proc. Natl. Acad. Sci. USA* **100**, 4372–4376 (2003).
- [5] Strano, E., Nicosia, V., Latora, V., Porta, S., & Barthélemy, M. Elementary processes governing the evolution of road networks. *Scientific reports* **2**, 296 (2012).
- [6] Singer, A. Zhao, Z., Shkolnisky, Y. & Hadani, R. Viewing angle classification of cryo-electron microscopy images using eigenvectors. *SIAM J. on Imaging Sci.* **4**(2), 723–759 (2011).
- [7] Singer, A. & Wu, H.-T. Two-dimensional tomography from noisy projections taken at unknown random directions. *SIAM J. on Imaging Sci.* **6**(1), 136–175 (2013).
- [8] Clauset, A., Moore, C. & Newman, M. E. J. Hierarchical structure and the prediction of missing links in networks. *Nature* **453**(7191), 98–101 (2008).
- [9] Guimerá, R. & Sales-Pardo, M. Missing and spurious interactions and the reconstruction of complex networks. *Proc. Nat. Acad. Sci.* **106**(52), 22073–22078 (2009).

- [10] Liben-Nowell, D. & Kleinberg, J. The link-prediction problem for social networks. *J. Amer. Soc. Info. Sci. Tech.* **58**(7), 1019–1031 (2007).
- [11] Bollobás, B. *Random Graphs, 2nd edition* (Cambridge University Press, Cambridge, 2001).
- [12] Newman, M. E. J. Models of the small world. *J. Stat. Phys.* **101**, 819–841 (2000).
- [13] Porter, M. A. Small-world network. *Scholarpedia* **7**(2), 1739 (2012).
- [14] Watts, D. J. & Strogatz, S. H. Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442 (1998).
- [15] Edelsbrunner, H. & Harer, J. Persistent homology – a survey. *Surv. Disc. Comp. Geom.*: **453**, 257–282 (2008).
- [16] Edelsbrunner, H. & Harer, J. *Computational Topology: An Introduction* (American Mathematical Society, 2010).
- [17] Carlsson, G. Topology and data. *Bull. Amer. Math. Soc.* **46**(2), 255–308 (2009).
- [18] Carlsson, G. & De Silva, V. Zigzag persistence. *Found. of Comput. Math* **10**(4), 367–405 (2010).
- [19] Dijkstra, E. W. A note on two problems in connexion with graphs. *Numer. Math.* **1**(1) 269–271 (1959).
- [20] Watts, D. J. A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci. USA* **99**(9), 5766–5771 (2002).
- [21] Porter M. A., Gleeson J. P. Dynamical systems on networks: A tutorial. Preprint at <http://arxiv.org/abs/1403.7663> (2014).
- [22] Cox, T. F. & Cox, M. A. *Multidimensional Scaling* (CRC Press, 2010).
- [23] Tenenbaum, J. B., de Silva, V. & Langford, J. C. A global geometric framework for nonlinear dimensionality reduction. *Science* **290**, 2319–2323 (2000).
- [24] Borsuk, K. On the imbedding of systems of compacta in simplicial complexes. *Fund. Math.*, 217–234 (1948).
- [25] Mischaikow, K. & Nanda, V. Morse theory for filtrations and efficient computation of persistent homology. *Discr. Comp. Geom.* **50**(2), 330–353 (2013).
- [26] Tausz, A., Vejdemo-Johansson, M. & Adams, H. JavaPlex: A research software package for persistent (co)homology. *Proceedings of ICMS 2014: Lecture Notes in Computer Science* 8592, 129–136 (2014).
- [27] Belkin, M. & Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comp.* **15**(6), 1373–1396 (2003).
- [28] Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013).
- [29] Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci. USA* **103**(7), 2015–2020 (2006).

- [30] Hufnagel, L., Brockmann, D. & Geisel, T. Forecast and control of epidemics in a globalized world. *Proc. Natl. Acad. Sci. USA* **101**, 15124 (2004).
- [31] He, D., Lui, R., Tse, C. K., Lin, X. & Stone, L. Global spatio-temporal patterns of influenza in the post-pandemic era. Preprint at <http://arxiv.org/abs/1407.5505> (2014).
- [32] Centola, D. The spread of behavior in an online social network experiment. *Science* **329**, 1194 (2010).
- [33] Centola, D. & Macy, M. Complex contagions and the weakness of long ties. *Amer. J. Sociol.* **113**(3), 702–734 (2007).
- [34] Centola, D., Eguíluz, V. M. & Macy, M. Cascade dynamics of complex propagation. *Physica A* **374**(1), 449–456 (2007).
- [35] Ghasemiefteh, G., Ebrahimi, R. & Gao, J. Complex contagion and the weakness of long ties in social networks: Revisited. In *Proc. Fourteenth ACM Conf. Elec. Commerce*, 507–524 (2013).
- [36] Kempe, D., Kleinberg, J. & Tardos, E. Maximizing the spread of influence through a social network. *Proc. Ninth ACM SIGKDD Inter. Conf. Knowl. Disc. Data Min.*, 137–146 (2003).
- [37] Pastor-Satorras, R. & Vespignani, A. Immunization of complex networks. *Phys. Rev. E* **65**(3), 036104 (2002).
- [38] Rhodes, C. J. & Anderson, R. M. Epidemic thresholds and vaccination in a lattice model of disease spread. *Theor. Pop. Bio.* **52**(2), 101–118 (1997).
- [39] Gleeson, J. P. Binary-state dynamics on complex networks: Pair approximation and beyond. *Phys. Rev. X* **3**: 021004 (2013).
- [40] Melnik, S., Ward, J. A., Gleeson, J. P., Porter, M. A. Multi-stage complex contagions. *Chaos* **23**, 013124 (2013).
- [41] Jeub, L. G. S., Balachandran, P., Porter, M. A., Mucha, P. J. & Mahoney, M. W. Think locally, act locally: Detection of small, medium-sized, and large communities in large networks. *Phys. Rev. E* **91**, 012821 (2015).
- [42] Petri, G., Scolamiero, M., Donato, I. & Vaccarino, F. Networks and Cycles: A Persistent Homology Approach to Complex Networks. In *Proc. Euro. Conf. Cplx. Sys.* 93–99 (2012).
- [43] Petri, G., Scolamiero, M., Donato, I. & Vaccarino, F. Topological strata of weighted complex networks. *PloS one* **8**(6), e66506 (2013).
- [44] Kahle, M. Topology of random simplicial complexes: A Survey. Preprint at <http://arxiv.org/abs/1301.7165> (2013).
- [45] Bobrowski, O. & Kahle, M. Topology of random geometric complexes: A Survey. Preprint at <http://arxiv.org/abs/1409.4734> (2014).
- [46] M. Boguñá, M., Papadopoulos, F. & Krioukov, D. Sustaining the Internet with hyperbolic mapping. *Nature Comms.* **1**, 62 (2010).

- [47] Serrano, M. A., Krioukov, D. & Boguñá, M. Self-similarity of complex networks and hidden metric spaces. *Phys. Rev. Lett.* **100**(7), 078701 (2008).
- [48] Hoff, P. D., Raftery, A. E. & Handcock, M. S. Latent space approaches to social network analysis. *J. Amer. Stat. Assoc.* **97**(460), 1090–1098 (2002).
- [49] Bingham, E. & Mannila, H. Random projection in dimensionality reduction: applications to image and text data. *Proc. Seventh ACM SIGKDD Intl. Conf. Know. Disc. and Data Min.*, 245–250 (2001).
- [50] Coifman, R. R. *et al.* Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc. Natl. Acad. Sci. USA* **102**(21), 7426–7431 (2005).
- [51] Gerber, S., Tasdizen, T. & Whitaker, R. Robust non-linear dimensionality reduction using successive 1-dimensional Laplacian eigenmaps. *Proc. 24th Intern. Conf. Mach. Learn.*, 281–288 (2007).
- [52] Sorzano, C. O. S., Vargas, J. & Montano, A. P. A survey of dimensionality reduction techniques. Preprint at <http://arxiv.org/abs/1403.2877> (2014).
- [53] Lafon, S. & Lee, A. B. Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE Trans. Pattern Anal. Mach. Intel.* **28**(9), 1393–1403 (2006).
- [54] Ghosh, R. & Lerman, K. Rethinking centrality: The role of dynamical processes in social network analysis. *Disc. Cont. Dyn. Sys. Ser. B* **19**(5), 1355–1372 (2014).
- [55] Lerman, K. & Ghosh, R. Network structure, topology, and dynamics in generalized models of synchronization. *Phys. Rev. E* **86**, 026108 (2012).