Topological data analysis for investigating contagions on networks

Heather A Harrington

Research Fellow

Mathematical Institute
University of Oxford

Collaboration

This work was done in collaboration with

- Florian Klimm, University of Oxford, United Kingdom
- Miro Kramar, Rutgers University, USA
- Konstantin Mischaikow, Rutgers University, USA
- Peter J. Mucha, University of North Carolina at Chapel Hill, USA
- Mason A. Porter, University of Oxford, United Kingdom
- Dane Taylor, University of North Carolina at Chapel Hill, USA

This work is available on arxiv ID 1408.1168.
In press, *Nature Communications*.

Contagion spreading on networks

Social contagion

- Information diffusion
  (innovations, memes, marketing)
- Belief and opinion
  (voting, political views, civil unrest)
- Behavior and health

Epidemic contagion

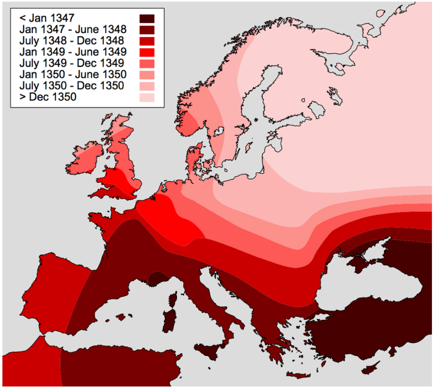
- Epidemiology for networks
  (social networks, technology)
- Preventing epidemics
  (immunization, malware, quarantine)

"Complex contagion"

- Adoption of a contagion requires
  multiple contacts with the contagion

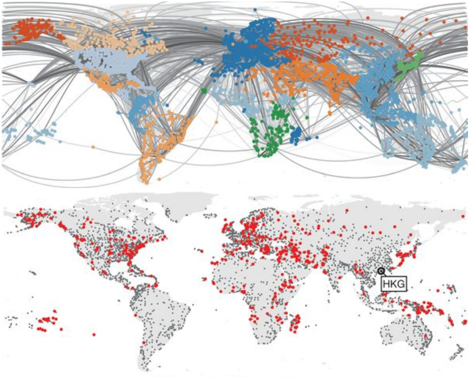
Epidemics on networks, then and now

Epidemics historically described by wave front propagation



Black death. Marvel et al (2014) arxiv 1310.2636

Modern epidemics driven airline network



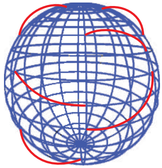
Brockman and Helbing (2013) Science

Noisy geometric networks

- Consider set $\mathcal{V}$ of network nodes with intrinsic locations $\{\mathbf{w}^{(i)}\}_{i\in\mathcal{V}}$ in a metric space (e.g., Earth's surface).
- Restrict to nodes that lie on a manifold $\mathcal{M}$ that is embedded in an ambient space $\mathcal{A}$ (i.e., $\mathbf{w}^{(i)} \in \mathcal{M} \subset \mathcal{A}$).
- "Node-to-node distance" refers to the distance between nodes in this embedding space $\mathcal{A}$ (here we use the Euclidean norm $\|\cdot\|_2$).

Place nodes in underlying manifold and add two types of two edge types:

- Geometric edges added between nearby nodes
- Non-geometric edges added uniformly at random

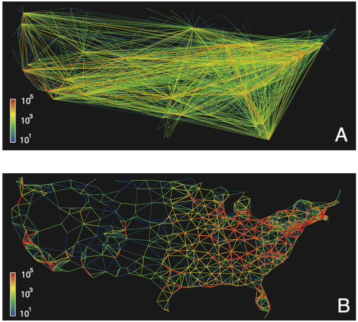


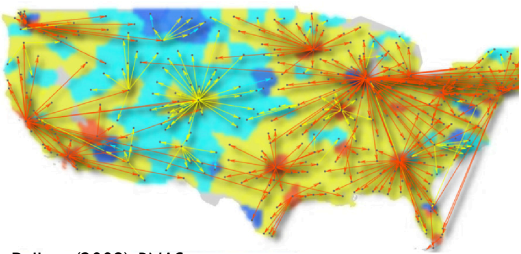
Origins

- Inherent heterogeneity in spatial networks
- Noise in networks/ noise in data

Does a contagion follow the underlying geometry?

- Network and geometry can disagree when long-range edges are present
- Will dynamics of a contagion follow the network's geometric embedding?
- For a network embedded on a manifold, to what extent does the manifold manifest in the dynamics?





- Balkan (2009) *PNAS*

Watts Threshold Model (WTM) for complex contagion

Watts (2002) PNAS

For time $t = 1, 2, \dots$ binary dynamics at each node $n \in \mathcal{V}$

- $x_n(t) = 1$ contagion adopted by time t
- $x_n(t) = 0$ contagion not adopted by time t

Node n adopts the contagion if the fraction $f_n(t)$ of its neighbors that have adopted the contagion surpasses a threshold T.

$$x_n(t+1) = 1 \text{ if } x_n(t) = 0 \text{ and } f_n(t) > T$$

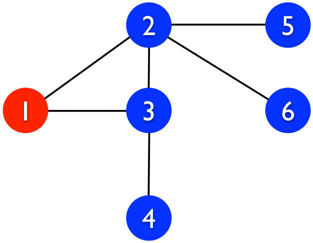
Otherwise $x_n(t+1) = x_n(t)$, i.e., no change

Note: the number of adopters in this model is non-decreasing.

Watts Threshold Model (WTM) for complex contagion

Watts (2002) PNAS

Example of WTM for $T = 0.3$



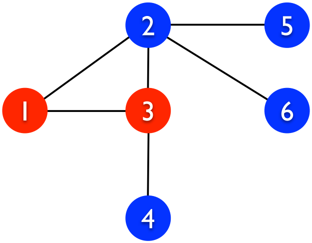
Timestep 0

Node	1	2	3	4	5	6
Activation time t	0					

Watts Threshold Model (WTM) for complex contagion

Watts (2002) PNAS

Example of WTM for T = 0.3



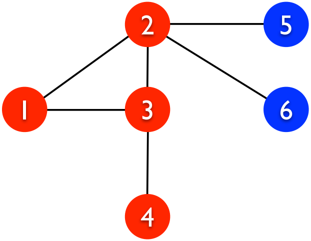
Timestep 1

Node	1	2	3	4	5	6
Activation time t	0		1			

Watts Threshold Model (WTM) for complex contagion

Watts (2002) PNAS

Example of WTM for $T = 0.3$



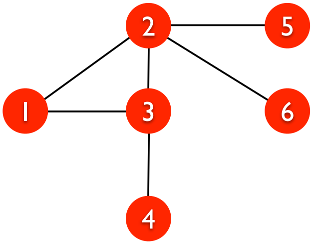
Timestep 2

Node	1	2	3	4	5	6
Activation time t	0	2	1	2		

Watts Threshold Model (WTM) for complex contagion

Watts (2002) PNAS

Example of WTM for $T = 0.3$



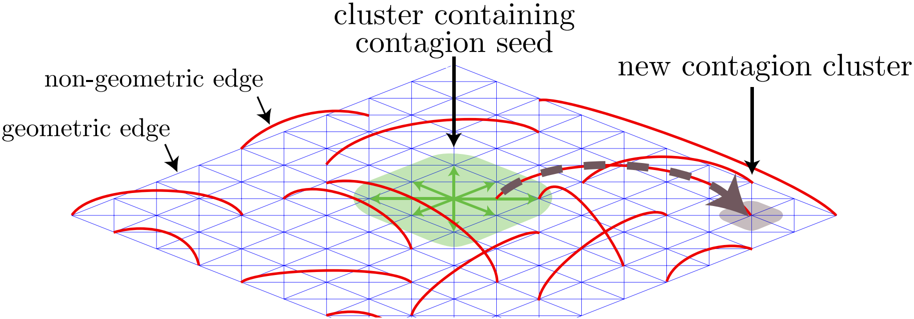
Timestep 3

Node	1	2	3	4	5	6
Activation time t	0	2	1	2	3	3

Contagion phenomena

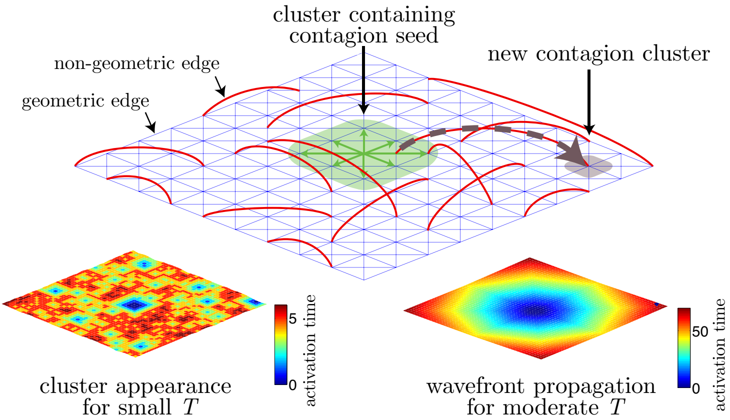
Spreading of contagion phenomena can be described by edge types:

- Wave front propagation (WFP) by spreading across geometric edges
- Appearance of new clusters (ANC) of contagion from spreading across non-geometric edges



WFP and APC depend on threshold T

Activation time is the time at which the node adopts the contagion



Noisy ring lattice

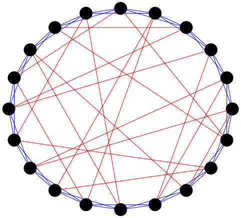
Aim I: Analyze the Watts threshold model (WTM) on noisy geometric network

Simple network model

- Ambient space: $\mathcal{A} = \mathbb{R}^2$
- Ring manifold: $\mathcal{M} = (x, y) \in \mathbb{R}^2$ s.t. $x^2 + y^2 = 1$
- Uniformly sampled a ring manifold

Network is given by 3 parameters:

1. Number of nodes N
2. Geometric degree d^G
3. Non-geometric degree d^{NG}
4. Ratio of non-geometric to geometric edges, $\alpha = d^{NG}/d^G$



In this example, $N = 20, d^G = 4, d^{NG} = 2, \alpha = 1/2$

Critical thresholds for noisy ring lattice

To what extent do the dynamics of a contagion spreads by WFP versus ANC?

Wavefront propagation (WFP) is governed by the critical thresholds:

$$T_k^{(\mathrm{WFP})} = \frac{d^{(\mathrm{G})}/2 - k}{d^{(\mathrm{G})} + d^{(\mathrm{NG})}}\,, \quad k = 0, 1 \ldots, d^{(\mathrm{G})}/2. \tag{1}$$

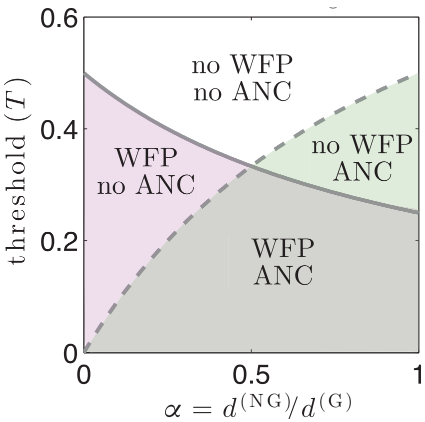
Note: for $T \geq T_0^{(\mathrm{WFP})}$, there is no WFP.

Appearance of new clusters (APC) has a sequence of critical thresholds:

$$T_k^{(\mathrm{ANC})} = \frac{d^{(\mathrm{NG})} - k}{d^{(\mathrm{G})} + d^{(\mathrm{NG})}}\,, \quad k = 0, 1 \ldots, d^{(\mathrm{NG})}. \tag{2}$$

Bifurcation diagram and trait regimes

Consider the case when $k = 0$



$$T_0^{(\mathrm{WFP})} = \frac{1}{(2+2\alpha)}$$

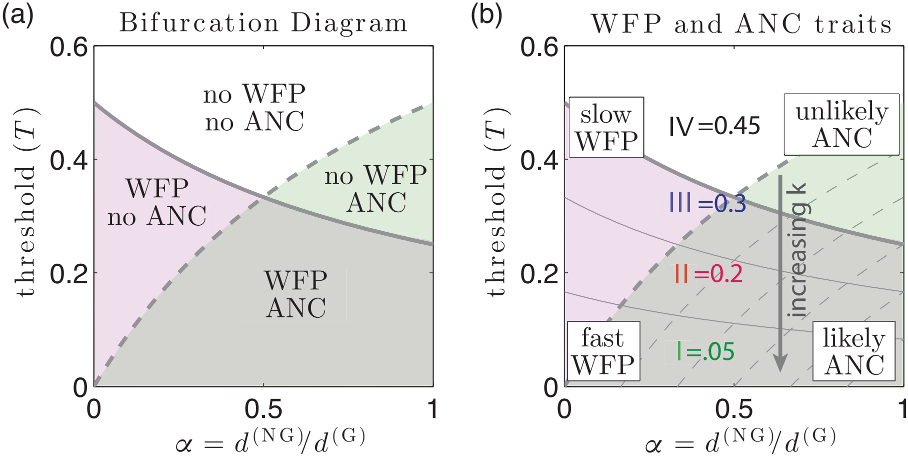
$$T_0^{(\mathrm{ANC})} = \frac{\alpha}{(\alpha+1)}$$

Lines intersect at $(\alpha, T) = (1/2, 1/3)$

Four regimes of contagion dynamics characterized by the presence versus absence of WFP and ANC.

Bifurcation diagram and trait regimes

Consider increasing values of k (decreasing lines) for fixed $d^G = 6$



Four regimes

$I: \ T \in (0; .125), \ II: \ T \in (.125; .25), \ III: \ T \in (.25; .375), \ IV: \ T > .375$

WTM maps: contagion maps based on WTM contagions

Aim II: To what extent do the spreading dynamics follow the manifold on which the network is embedded? To study this we construct and analyze WTM maps!

A WTM map is a nonlinear map of nodes to a high-dimensional point cloud in a metric space based on the activation times from N realisations of WTM contagions.

- Initialise the j-th contagion centred at node $j = 1, \ldots N$.
- Record activation time $x_j^{(i)}$ for each node i and contagion j.
- Each node i maps to the point $\mathbf{x}^{(i)} = [x_1^{(i)}, x_2^{(i)}, \ldots, x_N^{(i)}]$

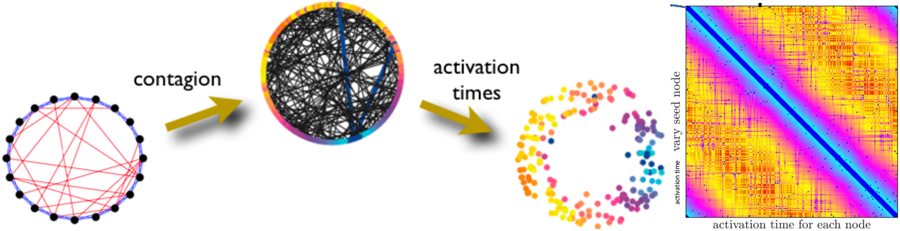
$$\mathcal{V} \mapsto \{\mathbf{x}^{(i)}\}_{i \in \mathcal{V}} \in \mathbb{R}^N$$

- The WTM map $\mathcal{V} \mapsto \{\mathbf{x}^{(i)}\}_{i \in \mathcal{V}} \in \mathbb{R}^N$ yields a high dimensional point cloud.

WTM maps: contagion maps based on WTM contagions

Aim II: To what extent do the spreading dynamics follow the manifold on which the network is embedded? To study this we construct and analyze WTM maps!

Visualise N dimensional WTM map after 2D mapping via PCA.

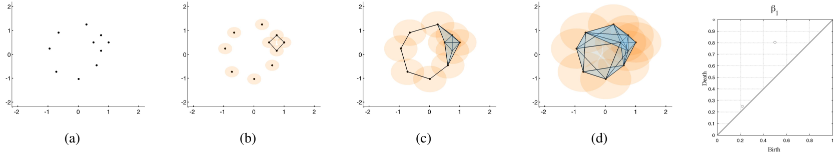


Observe ring structure already. For each node, have activation time under each possible contagion with different seed node. $N \times N$ non-symmetric matrix.

Analysis using geometry, topology and dimensionality

How does the distance between two nodes in a point cloud from a WTM map relate to the distance between those nodes in the original metric-space?

- **Geometry** of the the node-to-node distances for the two point clouds (WTM map and underlying manifold) is compared by computing Pearson correlation coefficient ρ.
- Persistence of 1-cycles (Δ describes lifespan of cycles) using **persistent homology**



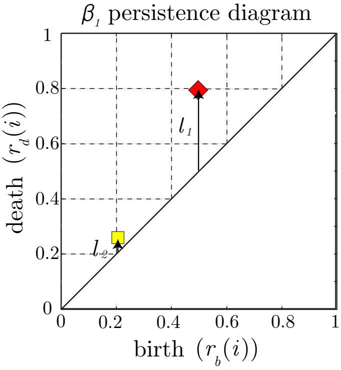
(a) (b) (c) (d)

- Embedding **dimension** P is computed by studying p-dimensional projections of the WTM map obtained via principle component analysis (such that the residual variance R_p for the projection onto $\mathbb{R}^p < 0.05$).

Analysis using topology: persistent homology

We are interested in persistence of 1-cycles

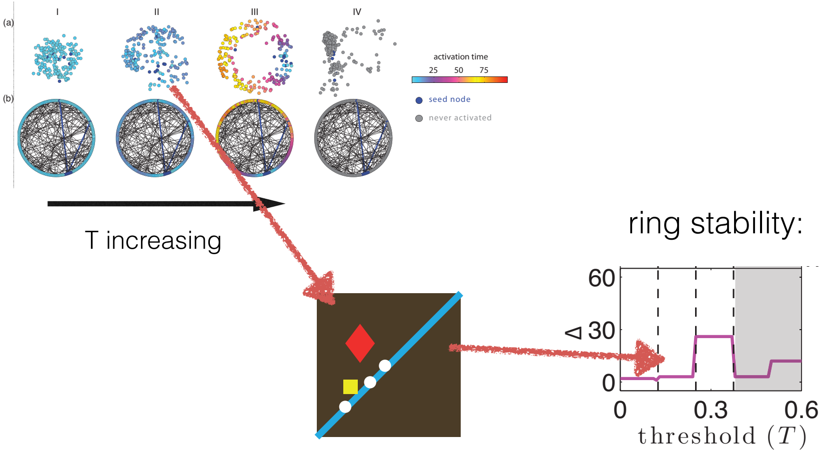
Analyze the point could by constructing a Vietoris-Rips filtration and calculate its persistent homology.



$\ell_i = r_d(i) - r_b(i)$
We define the ring stability $\Delta = \ell_1 - \ell_2$.

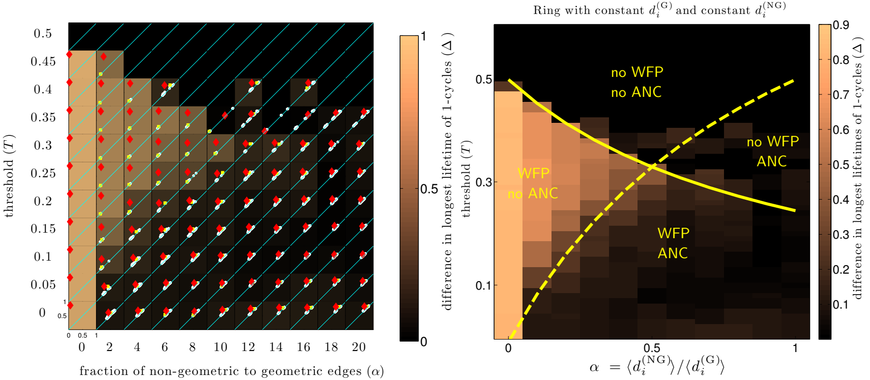
Analysis using topology: persistent homology

We are interested in persistence of 1-cycles



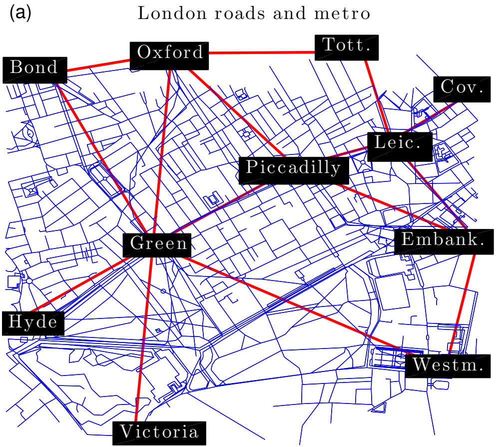
We define the ring stability $\Delta = \ell_1 - \ell_2$.

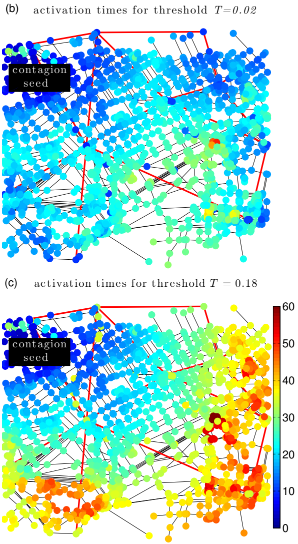
WTM maps analysis of noisy ring lattice



For a WFP dominant regime, WTM maps recover the topology, (as well as geometry, and dimensionality) of the network's underlying manifold even in the presence of non-geometric edges.

Application to London transit network

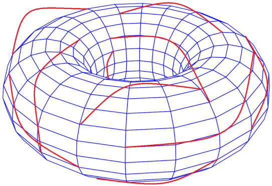




- Roads are geometric edges
- Underground stations are non-geometric
- Geometry is sensitive to threshold

Summary and outlook

- Studied WTM contagion on noisy geometric network and analyzed the spread of the contagion as parameters change. Classified WTM.
- Constructed WTM map which *map nodes as point cloud* based on several realisations of contagion on network.
- WTM map dynamics that are dominated by WFP recovers geometry, dimension and topology of underlying manifold.
- Applied WTM maps to London transit network and found agreement with moderate T.



Extending to other network geometries and contagion models. (Barbara Mahler)

Acknowledgements

This work is available on arxiv ID 1408.1168.
In press, Nature Communications.

Collaborators

- **Florian Klimm**, University of Oxford, United Kingdom
- Miro Kramar, Rutgers University, USA
- Konstantin Mischaikow, Rutgers University, USA
- Peter J. Mucha, University of North Carolina at Chapel Hill, USA
- Mason A. Porter, University of Oxford, United Kingdom
- **Dane Taylor**, University of North Carolina at Chapel Hill, USA

Thanks to Hal Schenck, Sayan Mukherjee, and Ezra Miller for helpful discussions.

Funding:

- King Abdullah University of Science and Technology (KAUST) KUK-C1-013-04
- SAMSI Low Dimensional Structure in High Dimensional Data workshop travel grant
- AMS Simons travel grant.