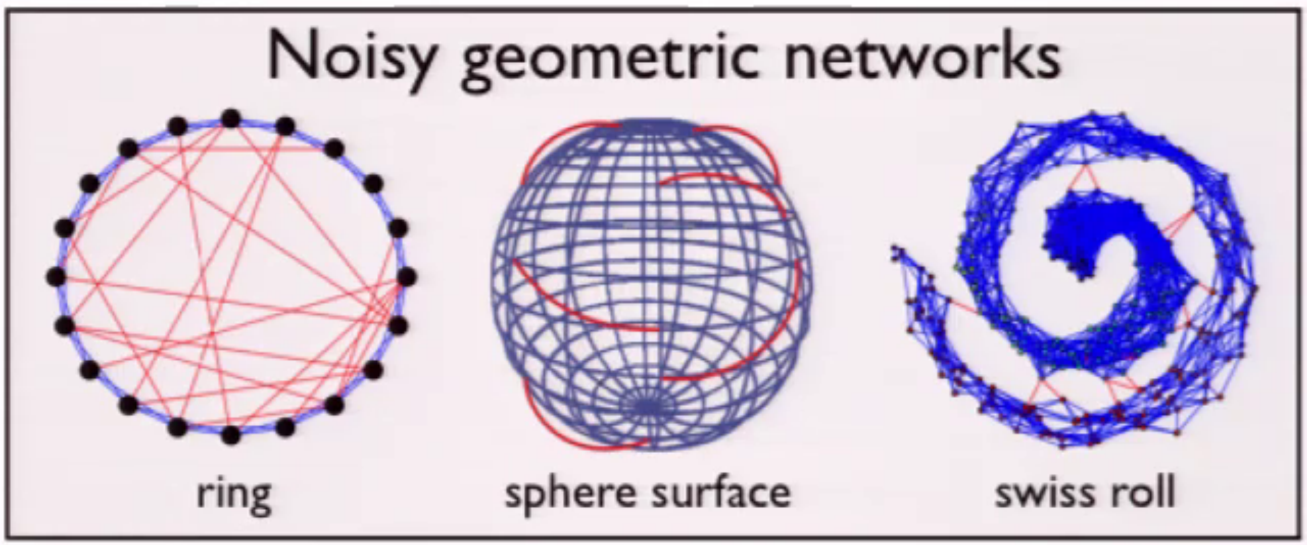
Contagions for topological data analysis of networks

preprint available at
arXiv:1408.1168



DANE TAYLOR, postdoctoral scholar
Department of Mathematics, University of North Carolina - Chapel Hill

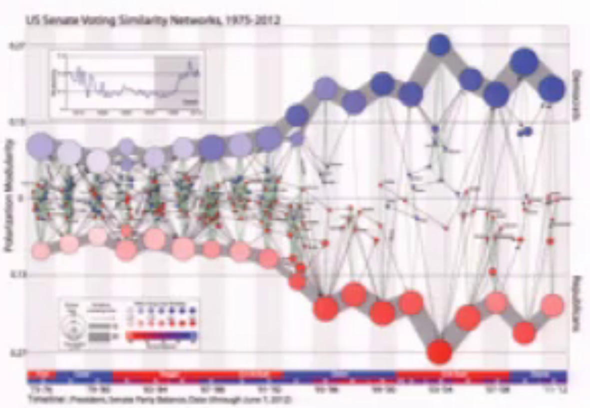
Dimension Reduction of Networks

- **Inferring *discrete* vs *continuous* structure**
  - Community detection gives *discrete* dimension reduction
  - Manifold learning gives *continuous* dimension reduction

Dimension Reduction of Networks

- Inferring *discrete* vs *continuous* structure
  - Community detection gives *discrete* dimension reduction
  - Manifold learning gives *continuous* dimension reduction

Party polarization



-Moody and Mucha, Network Science (2013)

Temporal-constrained edges

-Jeub *et al*, arXiv: 1403.3795 (2014)

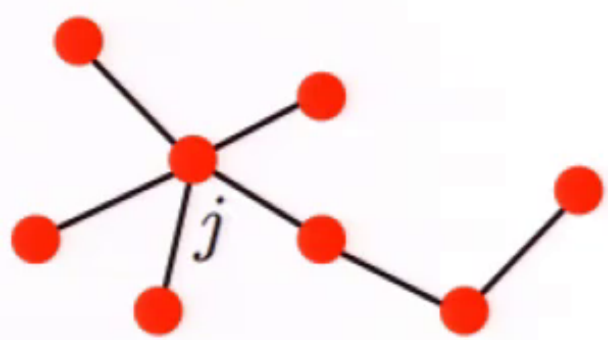
Low-dimensional structure in Senate data

Manifold Structure in Networks and Data

- **Previous research focusing on point cloud data**
  - Isomap (shortest paths)- Tenenbaum, de Silva & Langford, *Science*, 2000
  - Laplacian eigenmap (diffusion) - Belkin & Niyogi, *Neural Comp.*, 2003
  - Diffusion Maps (diffusion) - Coifman et al., *PNAS*, 2005
- **Contagion maps focusing on network data**
  - 2 applications: studying contagions, denoising networks

Dimension Reduction with Contagions - Topology

- Contagion-based analysis of networks from the perspective of *computational topology*
  - "Persistent homology – a survey," Edelsbrunner and Harer (2008)
- Monotonic, irreversible contagions yield a *filtrations* of a network



Set of infected nodes

$$j = I_j(0) \subseteq I_j(1) \subseteq I_j(2) \subseteq \cdots$$

a "filtration"

A Set of Filtrations can Induce a Metric

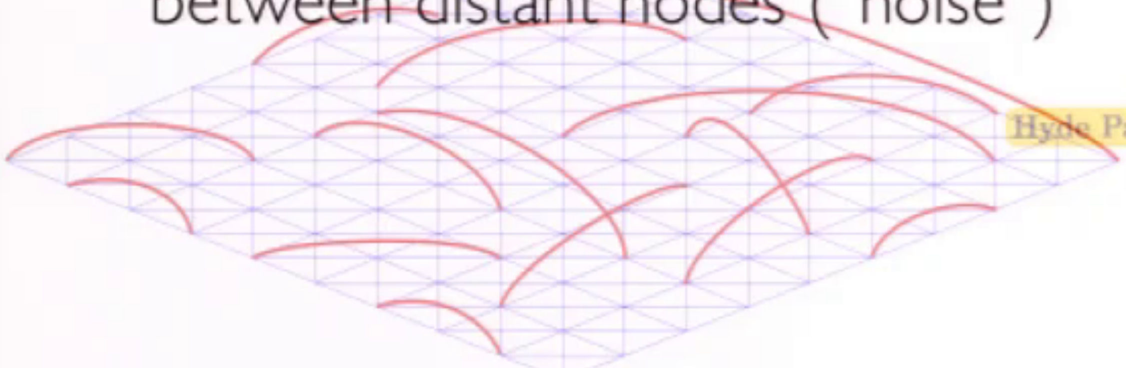
- **Theorem** (Filtration-Induced Metrics) -see our SI in *arXiv*: 1408.1168
  - Let $y_j^{(i)}$ denote the "activation time" at which node i adopts a contagion initialized on node j
  - The following is a metric on the set of nodes:

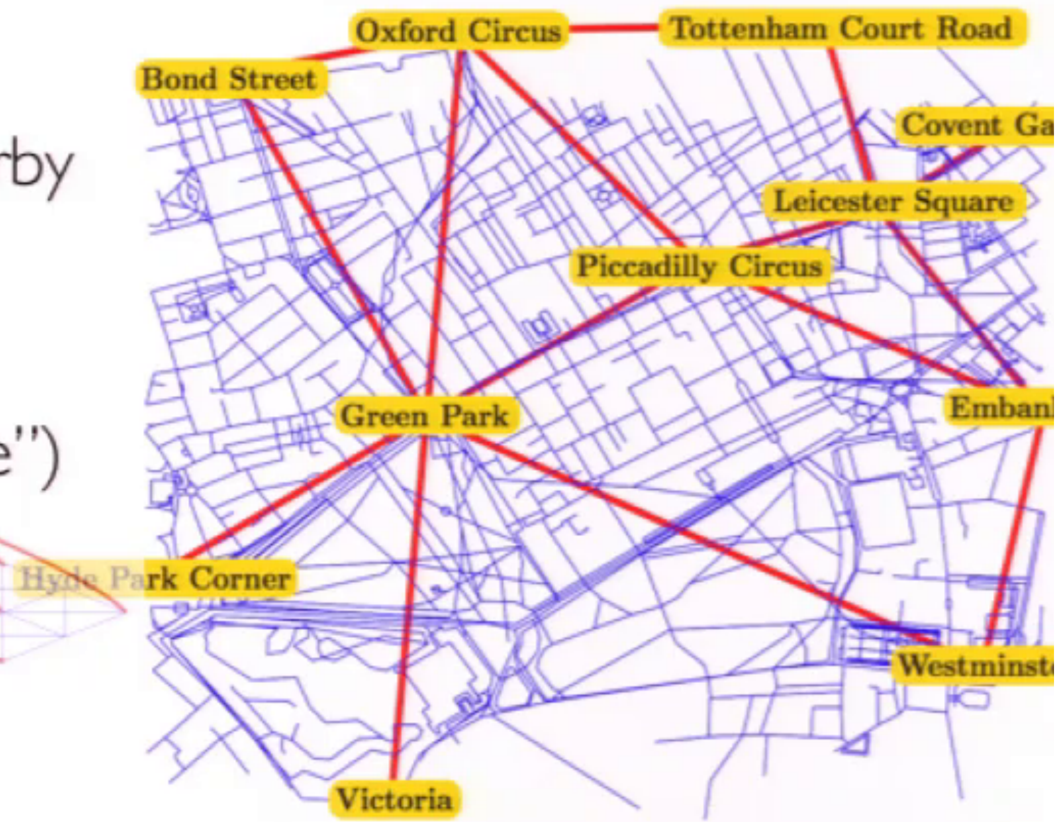
$$d(i,j) = y_j^{(i)} + y_i^{(j)}$$

- **Contagion transit gives a notion of "distance"**
- **Metrics derived from dynamics is not entirely new**
  - see "Diffusion Maps," RR Coifman et al (2005)

Dimension Reduction of Noisy Geometric Networks

- Nodes have intrinsic locations on a smooth manifold
- Two edge types:
  - Geometric edges added deterministically between nearby nodes
  - Non-geometric edges added between distant nodes ("noise")

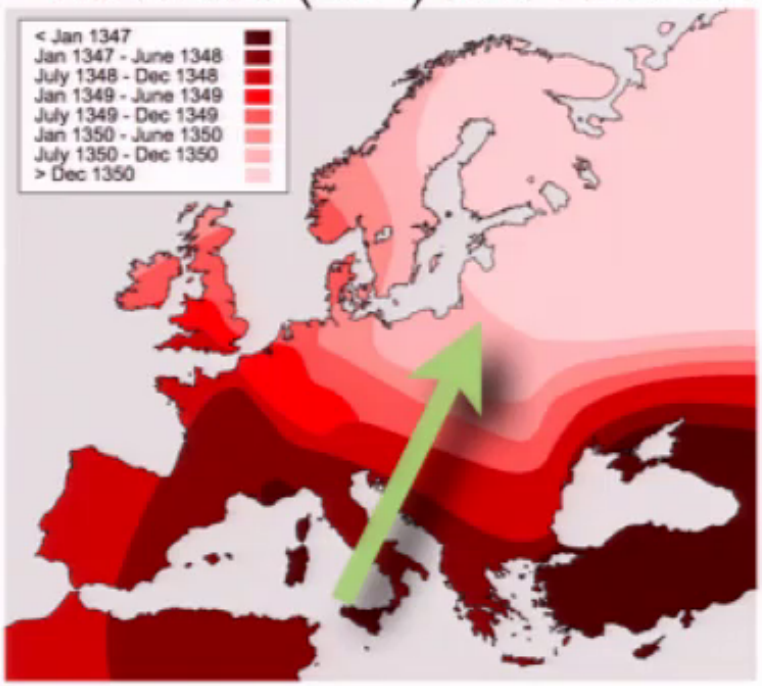




Contagions on Noisy Geometric Networks

- Contagions not completely understood

-Marvel et al (2014) *arXiv* 1310.2636

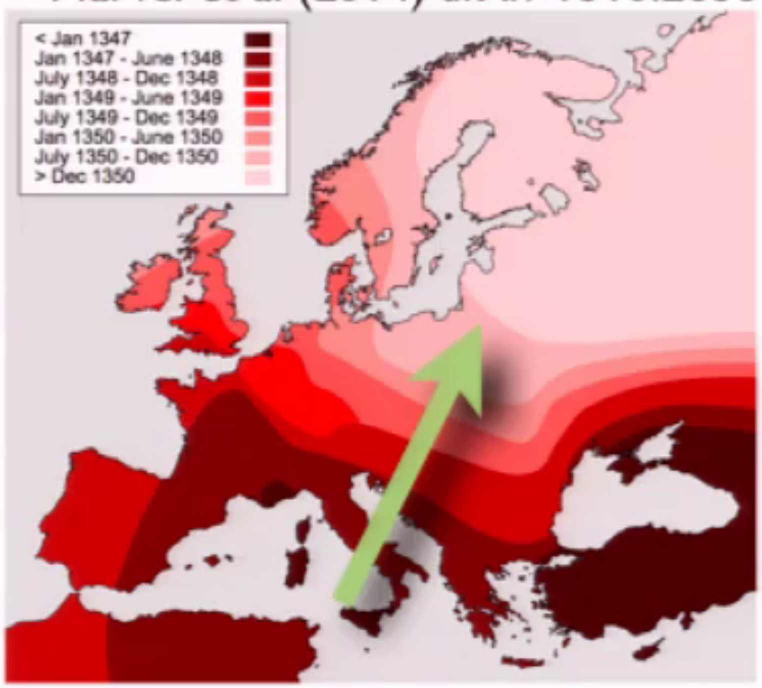


Epidemics historically spread by
wavefront propagation (WFP)

Contagions on Noisy Geometric Networks

- Contagions not completely understood

-Marvel et al (2014) *arXiv* 1310.2636



Epidemics historically spread by wavefront propagation (WFP)

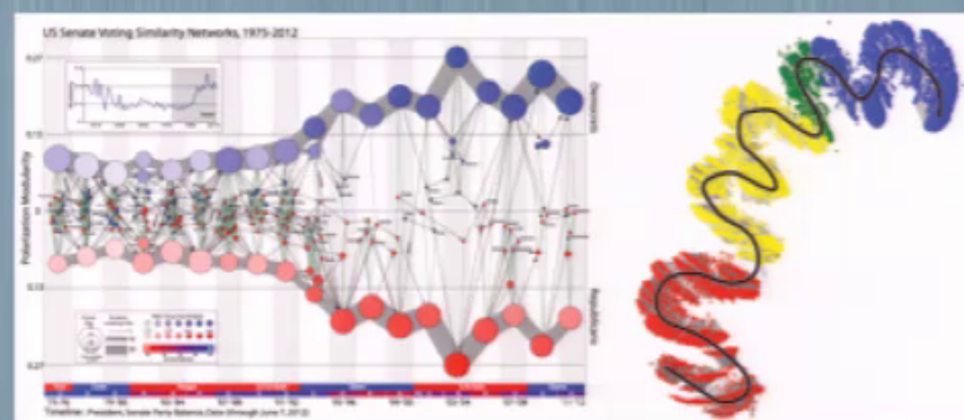
-Brockmann and Helbing (2013) *Science*

Modern epidemics dominated by the appearance of new clusters (ANC)

Dual Motivations

- Study contagions using an approach from high-dimensional data analysis

- Study low-dimensional (manifold) structure in networks using contagions



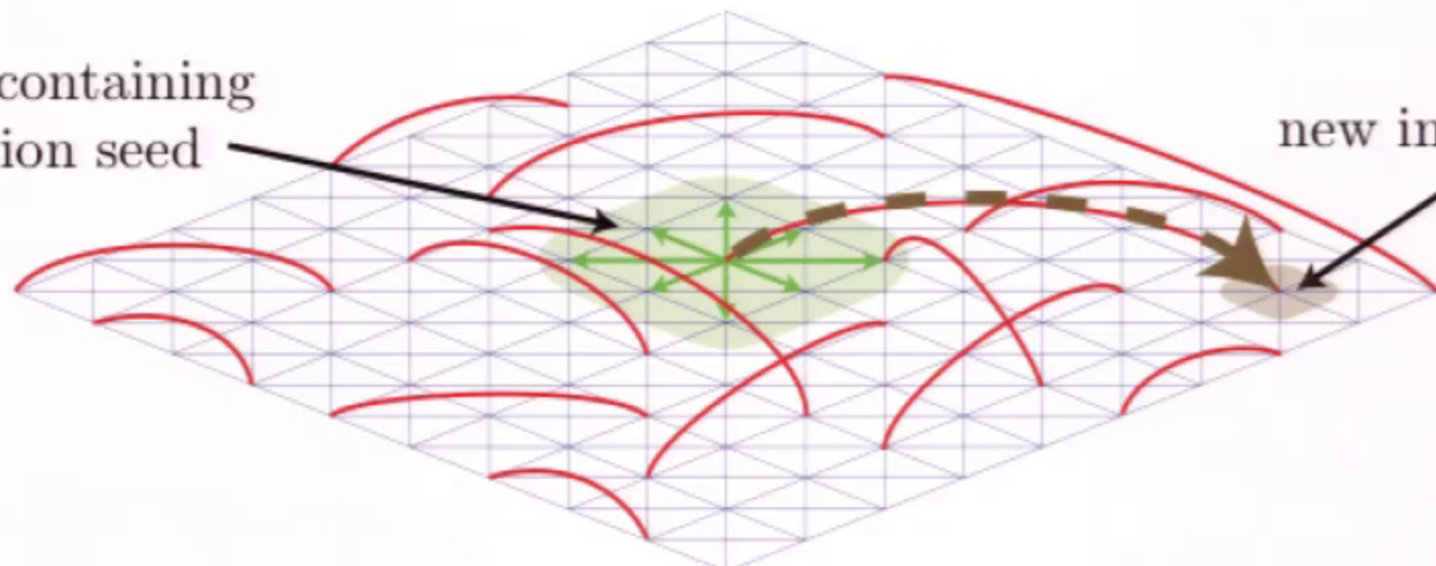
Watts Threshold Model (WTM) for Complex Contagions

- Discrete time, $t = 1, 2, \ldots$ binary state dynamics in which each node $n \in \mathcal{V}$ has one of two states:
  - $x_n(t) = 1$ indicates adopted contagion by time
  - $x_n(t) = 0$ indicates non-adoption
- Let $f_n(t)$ denote the fraction of infected neighbors for node n

Watts Threshold Model (WTM) for Complex Contagions

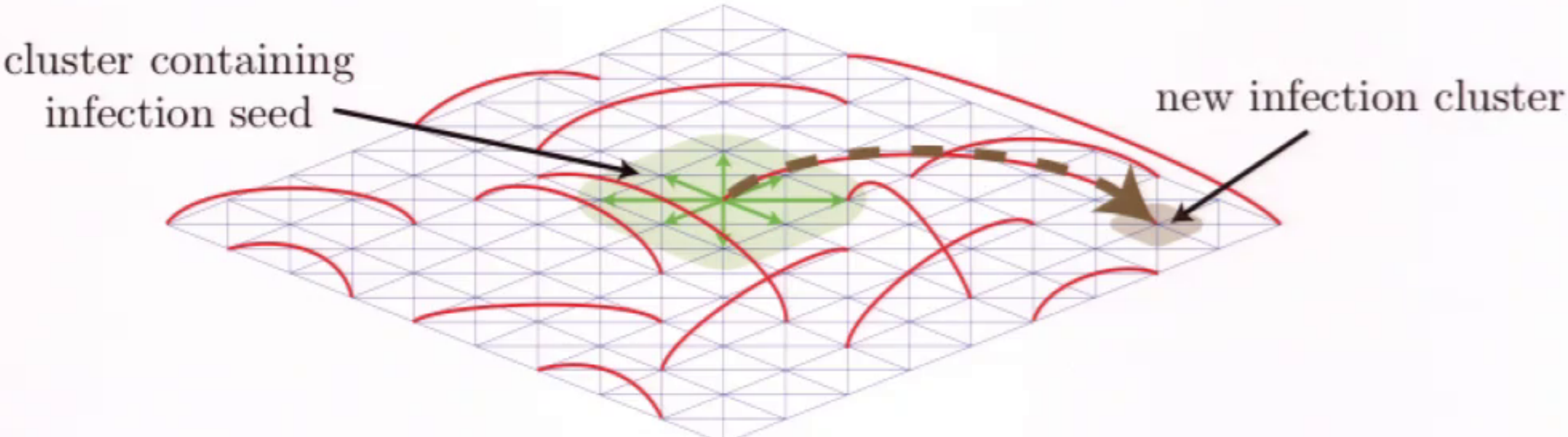
- Discrete time, $t = 1, 2, \ldots$ binary state dynamics in which each node $n \in \mathcal{V}$ has one of two states:
  - $x_n(t) = 1$ indicates adopted contagion by time
  - $x_n(t) = 0$ indicates non-adoption
- Let $f_n(t)$ denote the fraction of infected neighbors for node n
- Node n will adopt the contagion upon the next time step if $f_n(t)$ surpasses a uniform threshold $f_n(t) > T$
- Adopting the contagion is an irreversible event
  - Gives rise to a "filtration" of the network!

WTM Contagions on Noisy Geometric Networks



- Wavefront propagation (WFP) by spreading across geometric edges
- The appearance of new clusters (ANC) of contagion from spreading across non-geometric edges

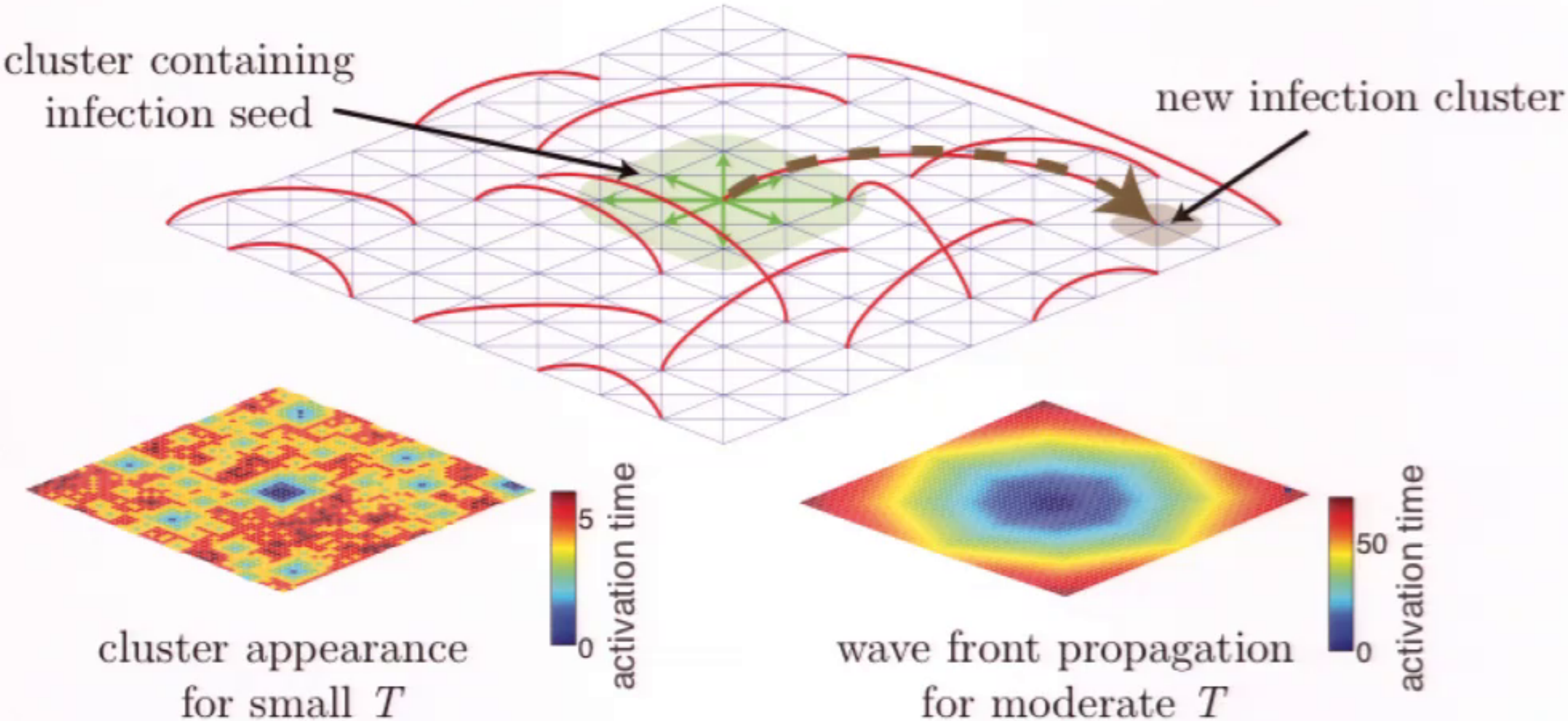
WTM Contagions on Noisy Geometric Networks



- WFP and ANC depend sensitively on T
  - Observable through node "activation times"

    -time at which a node adopts the contagion

WTM Contagions on Noisy Geometric Networks

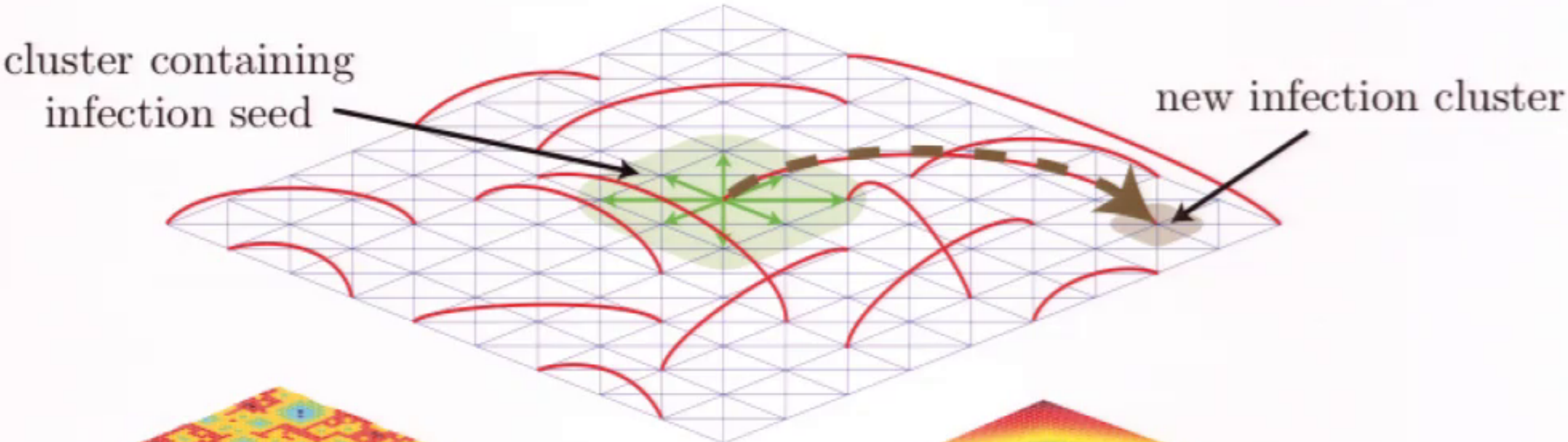


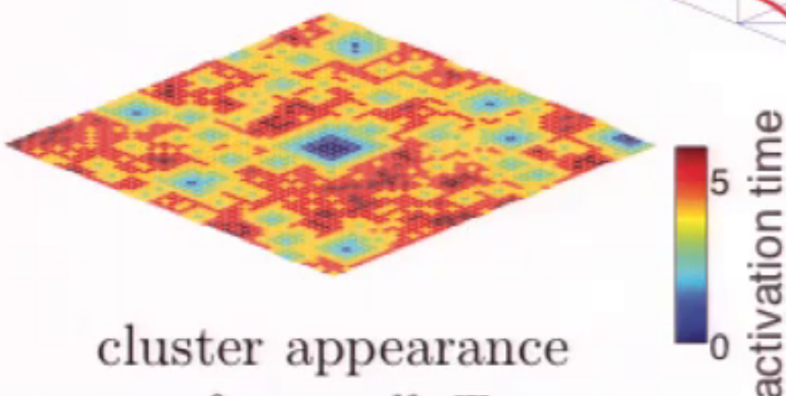
Activation Times Across Many Contagions

- Consider activation times for many contagions on a network
- Initialize the j-th contagion centered at node $j = 1, \ldots, N$
- Record activation time $y_j^{(i)}$ for each node i and contagion j
- Define vector of activation times $\mathbf{y}^{(i)} = \left[y_1^{(i)}, \ldots, y_N^{(i)}\right]$
- This defines a "WTM map"

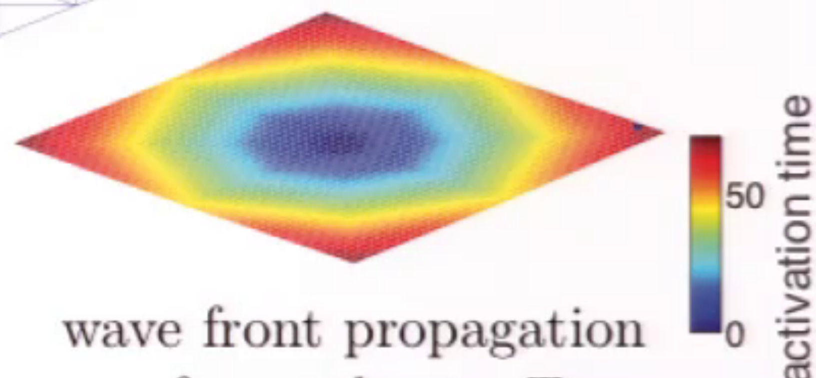
$$\mathcal{V} \mapsto \{\mathbf{y}^{(i)}\}_{i=1}^{N} \in \mathbb{R}^N$$

WTM Contagions on Noisy Geometric Networks





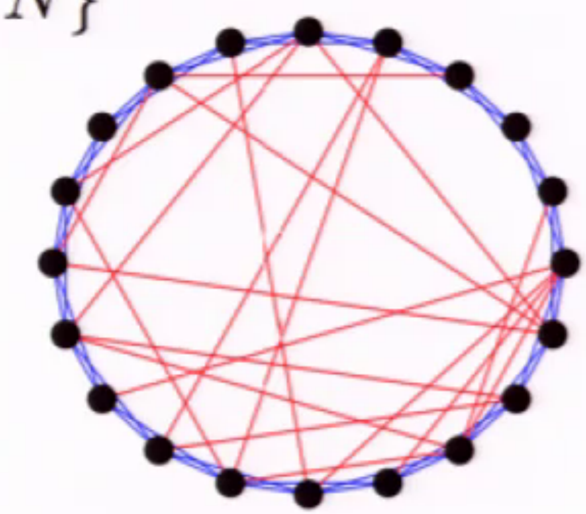
cluster appearance
for small T



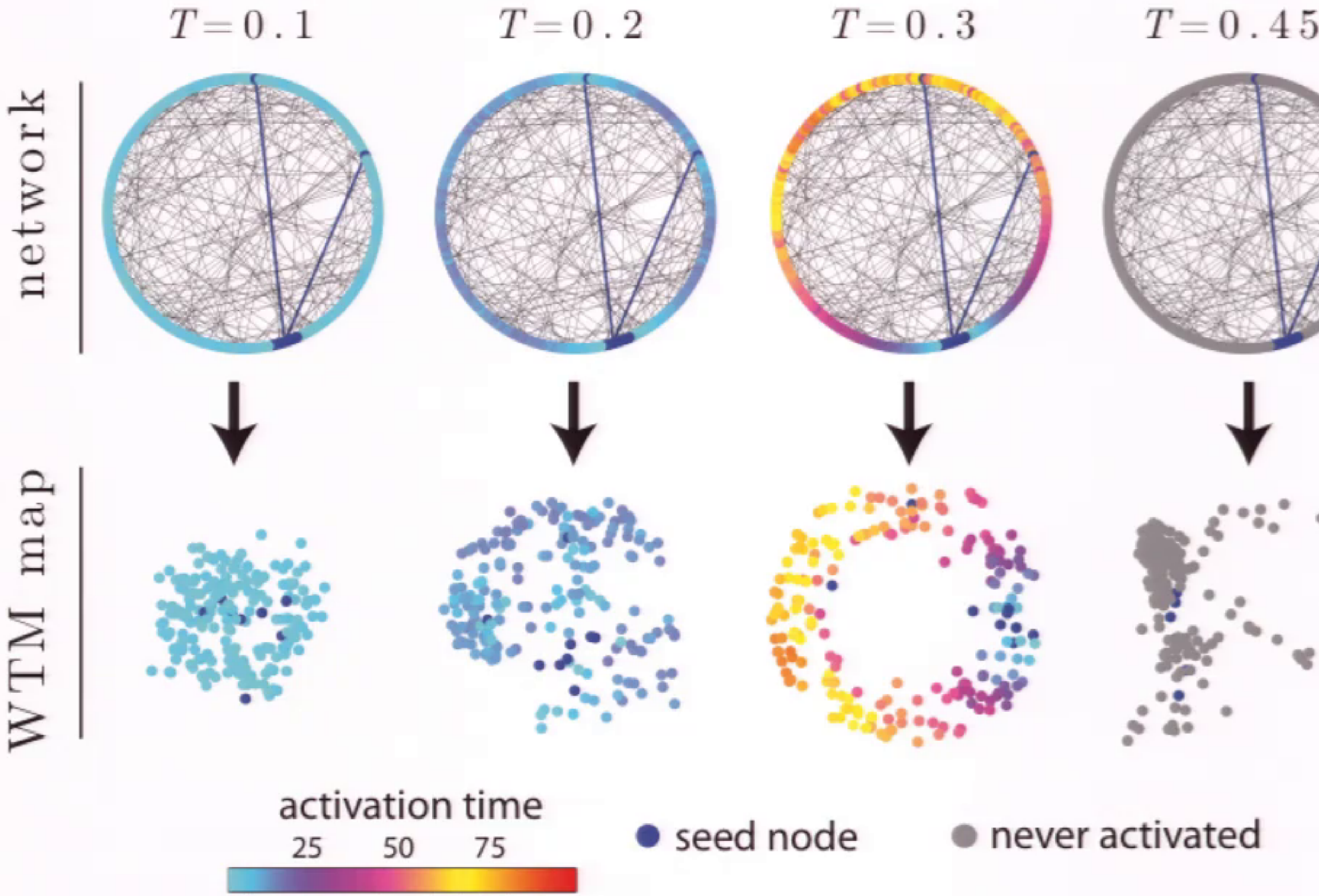
wave front propagation
for moderate T

Experiment with a Noisy Ring Lattice

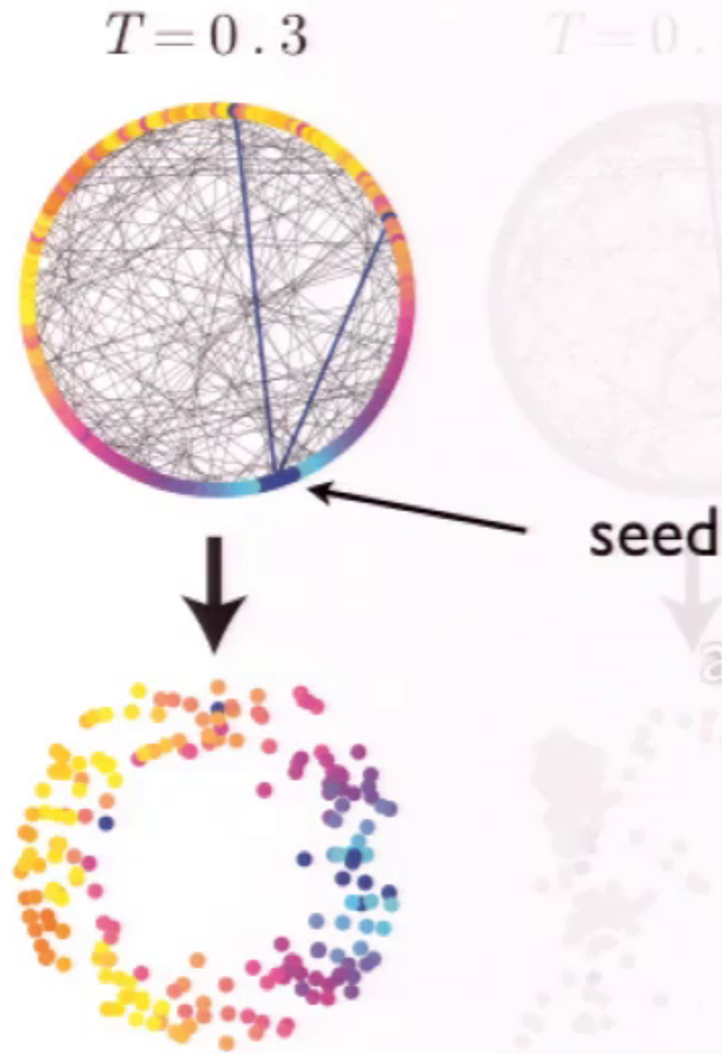
- Examine WTM maps with several thresholds T applied to a noisy geometric network embedded on the unit circle
- Network given by 3 parameters:
  - N is the number of nodes, $\mathcal{V} = \{1, 2, \ldots, N\}$
  - d^G is the geometric degree
  - d^{NG} is the non-geometric degree
  - $\alpha = d^{NG}/d^G$ is the ratio of non-geometric to geometric edges



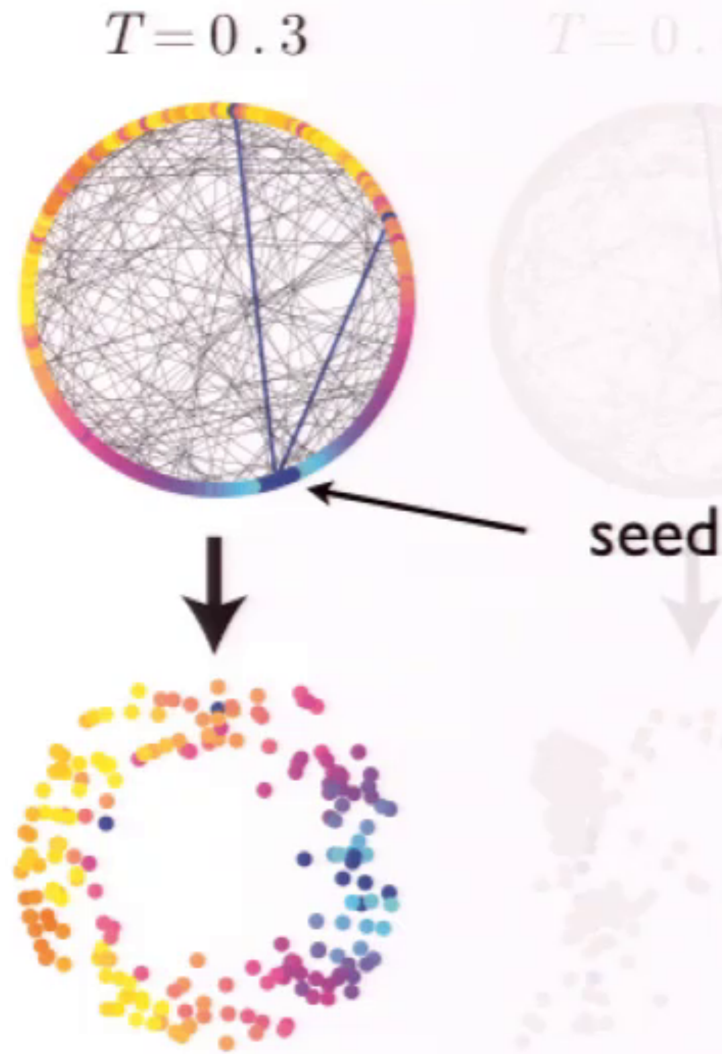
WTM maps for several thresholds



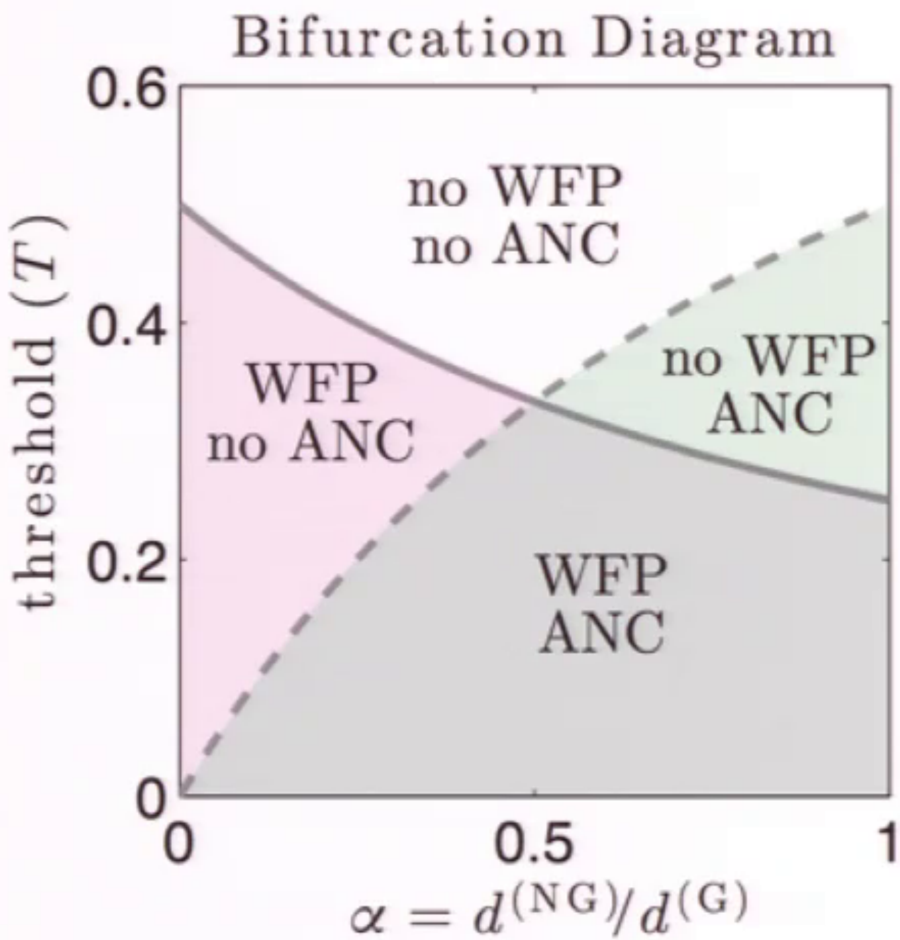
- Consider WTM map corresponding to contagions exhibiting WFP and no ANC
- The ring manifold underlying the noisy geometric network appears in the point cloud



- Consider WTM map corresponding to contagions exhibiting WFP and no ANC
- The ring manifold underlying the noisy geometric network appears in the point cloud
- Study point cloud to infer manifold structure in the network
- Ring presence indicates that WFP dominates ANC



Guide Study with Analysis of WTM Contagions

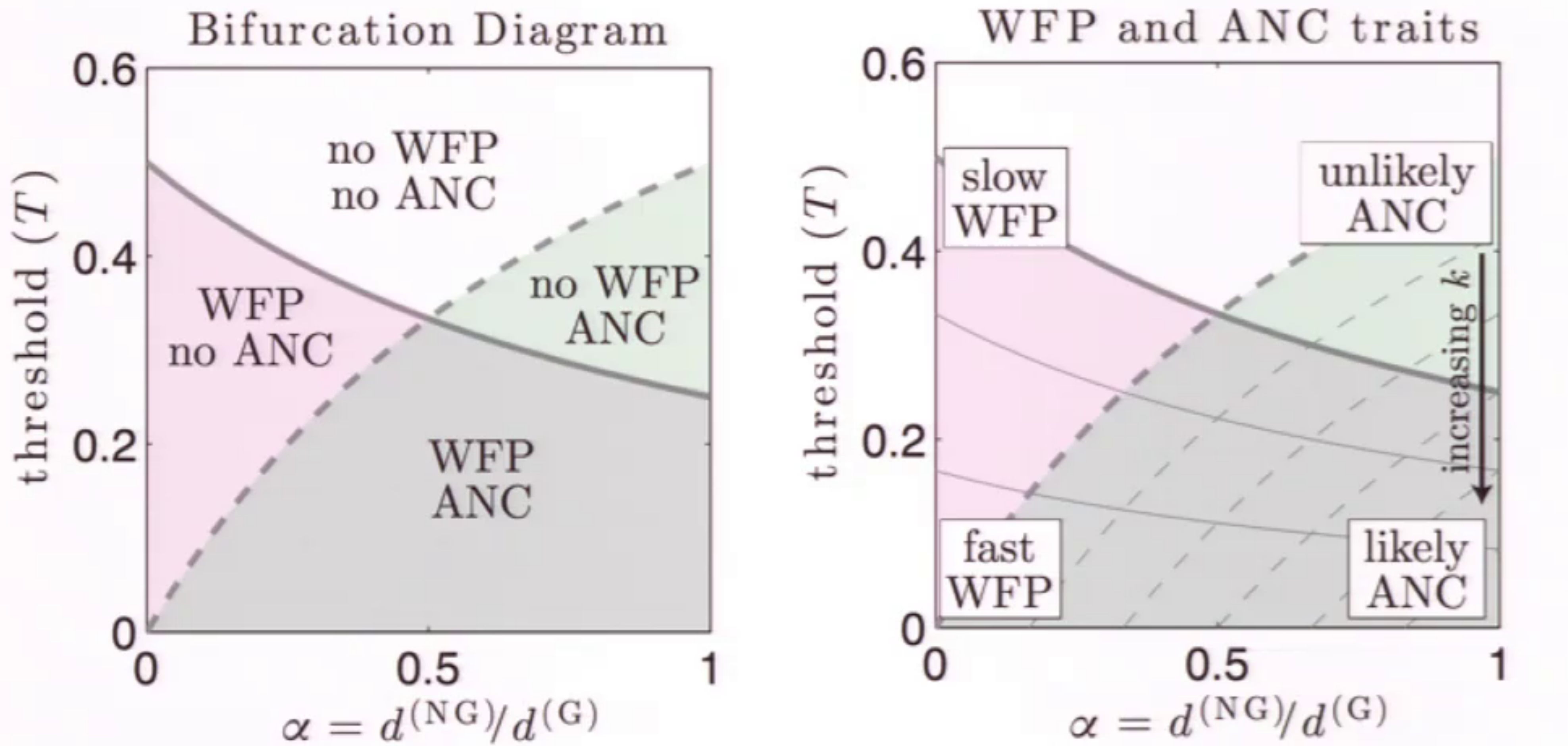


- **We analyze WTM contagions for the limit of large networks**
  - Critical threshold values determine absence/presence of WFP and ANC

—— $T_0^{WFP} = \dfrac{1}{2 + 2\alpha}$

- - - $T_0^{ANC} = \dfrac{\alpha}{\alpha + 1}$

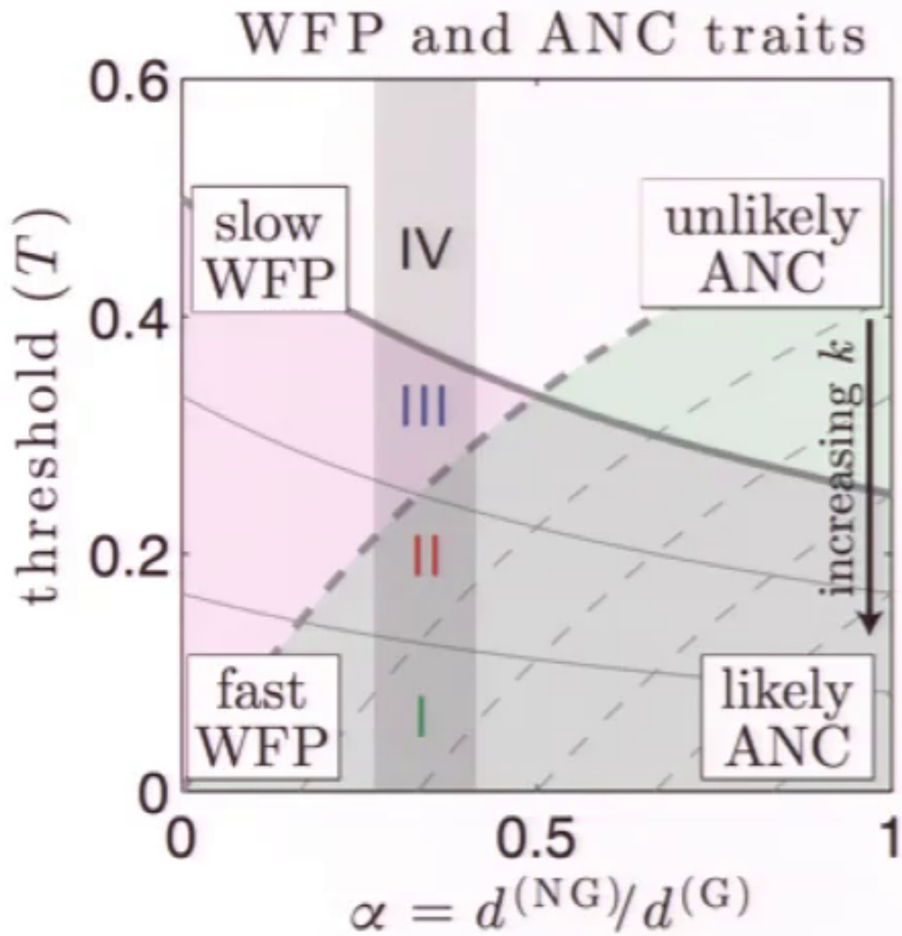
Guide Study with Analysis of WTM Contagions



- results shown for $d^G = 6$

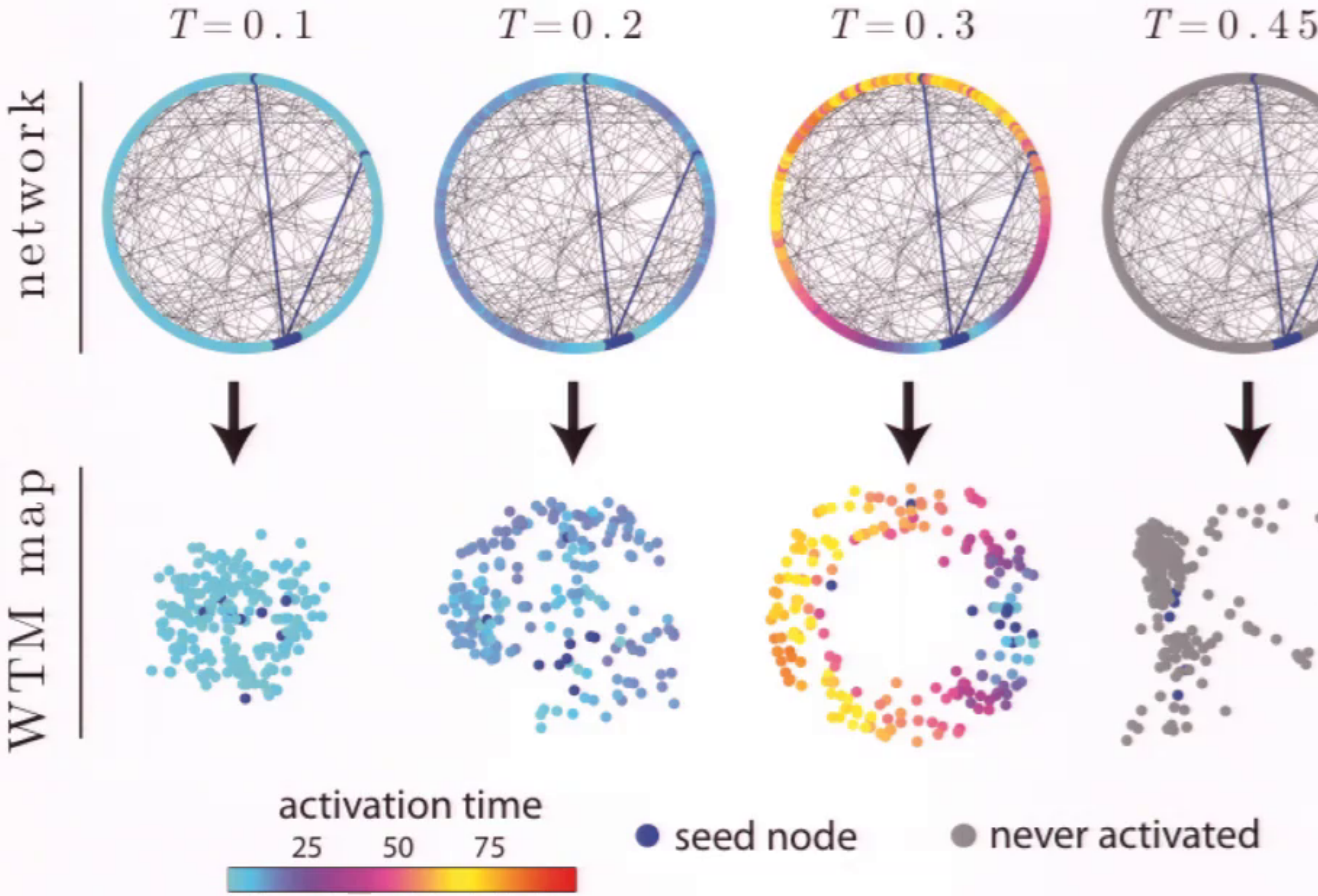
Guide Study with Analysis of WTM Contagions

- Fixing $\alpha = 1/3$, we observe 4 regimes
- These regimes guided our choice of thresholds for the experiment

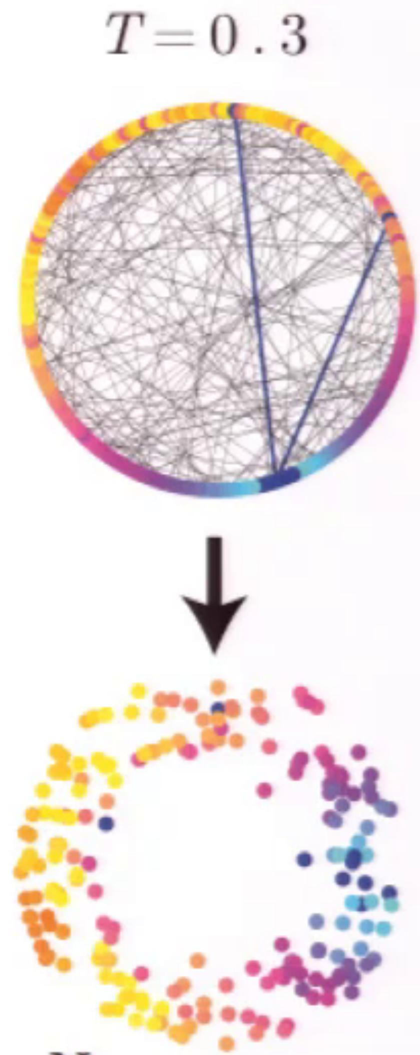


- results shown for $d^G = 6$

WTM maps for several thresholds



Noisy Ring Lattice



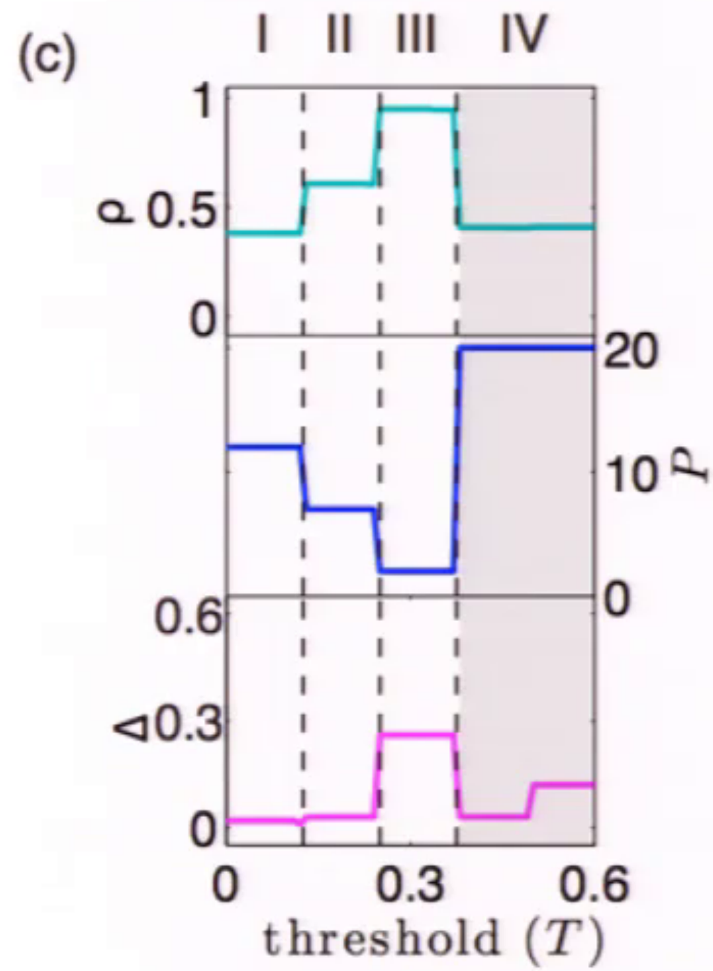
- We want to compare the ring manifold to the point cloud structure QUANTITATIVELY.

$$\mathcal{V} \mapsto \{\mathbf{y}^{(i)}\}_{i=1}^{N} \in \mathbb{R}^N$$

Study 3 Manifold Properties

- Geometry via ρ
- Dimensionality via P
- Topology via Δ

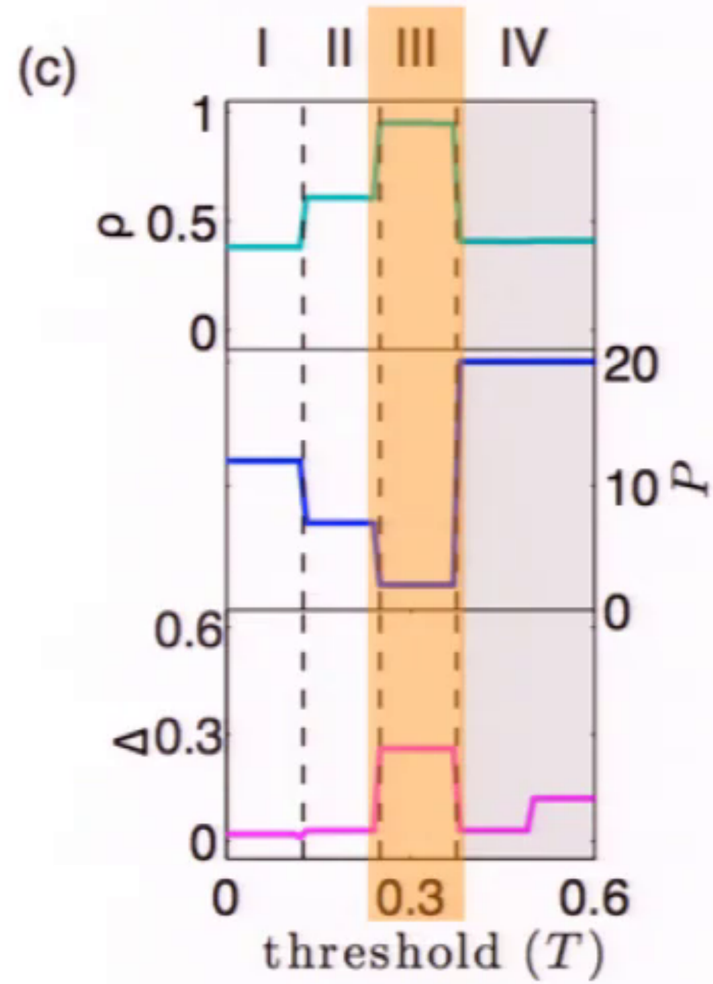
- Subsequently, we can study these properties for networks using contagions



Study 3 Manifold Properties

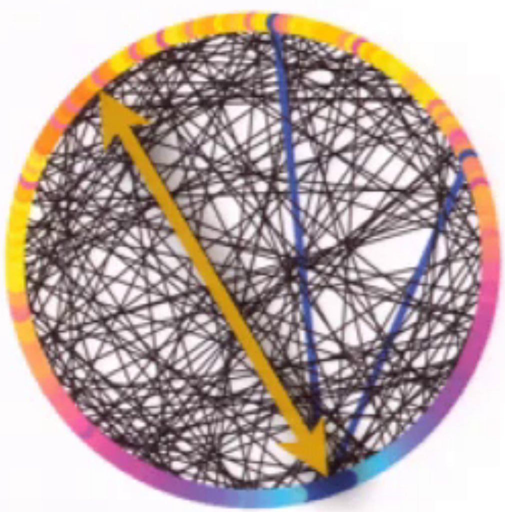
- Geometry via ρ
- Dimensionality via P
- Topology via Δ

- Subsequently, we can study these properties for networks using contagions



Point Cloud Analysis

- **Geometry** is examined via a *correlation coefficient* ρ that compares distances between node locations and distances between points

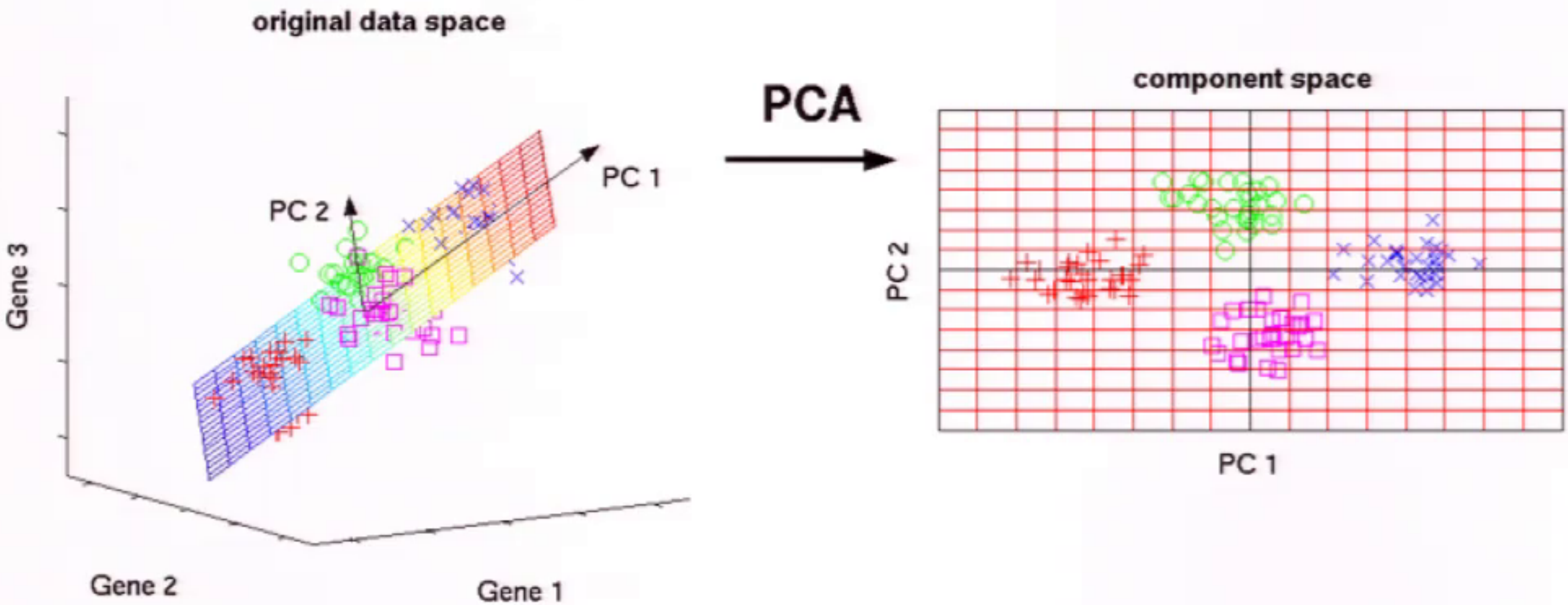


network

WTM map

Point Cloud Analysis

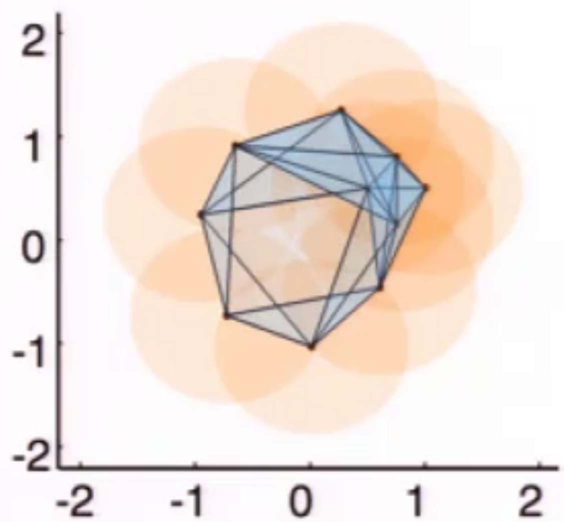
- **Embedding dimension** P is given by the *residual variance*
  - Dimension such that Principle Component Analysis (PCA) retains 95% of the point cloud's variance

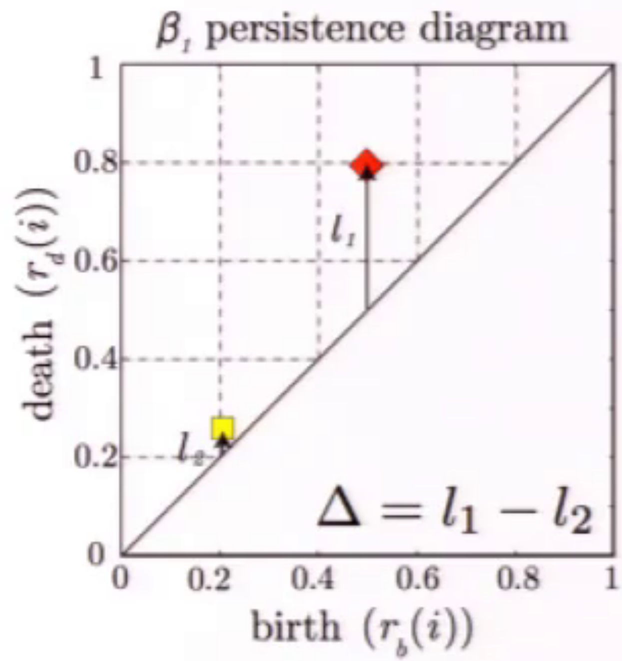


Point Cloud Analysis

- **Topology** of the ring measured by examining the *persistent homology* in the point cloud using a Viteoris-Rips Filtration

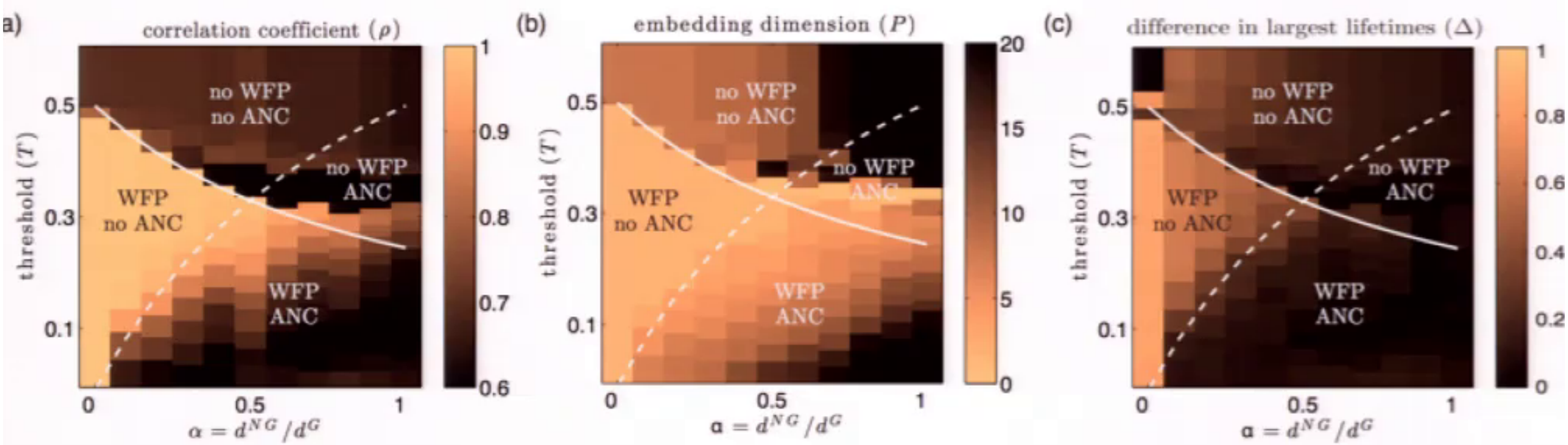
"Persistent homology – a survey," Edelsbrunner and Harer (2008)





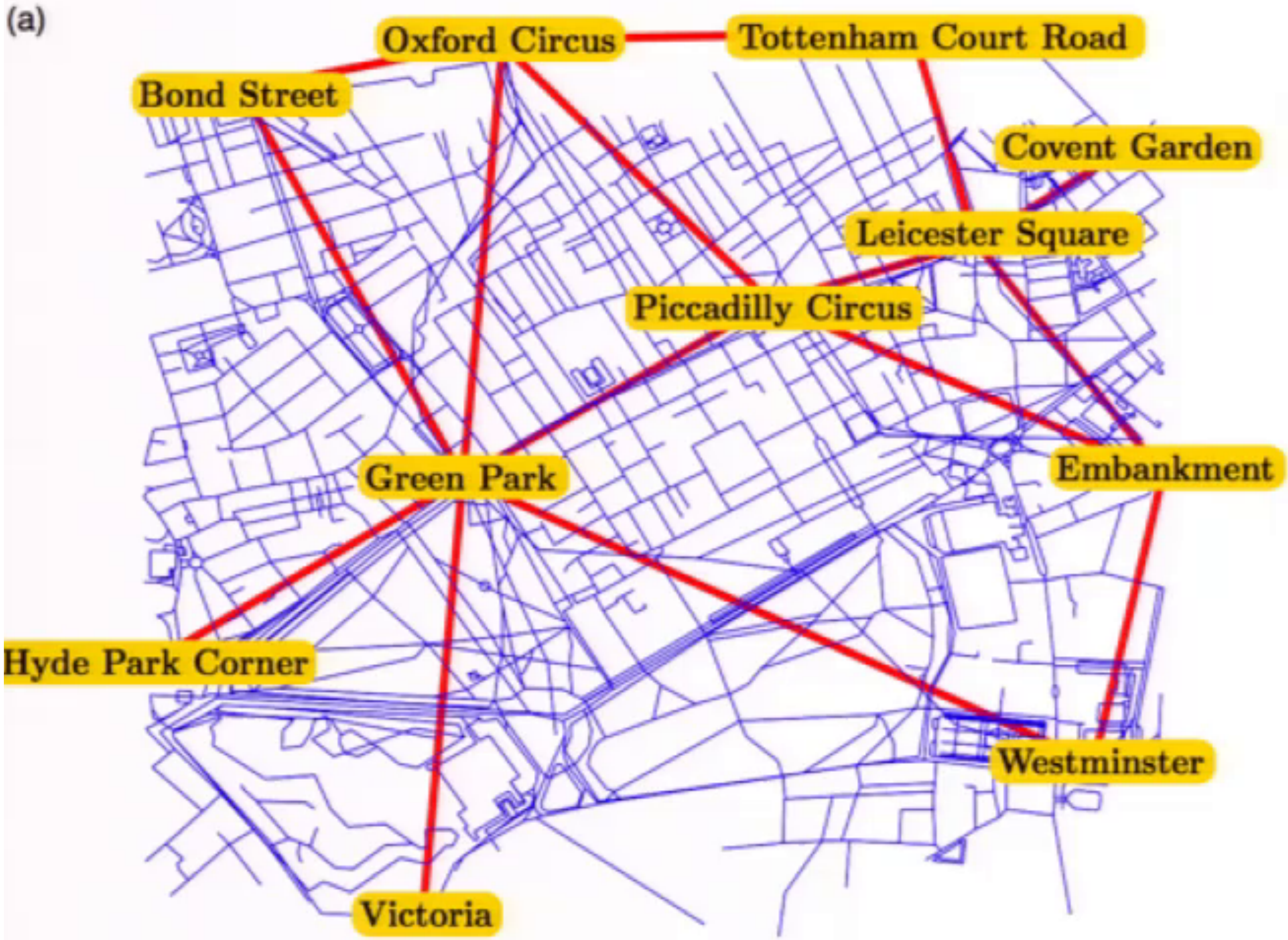
Comparing WTM Maps to Bifurcation Analysis

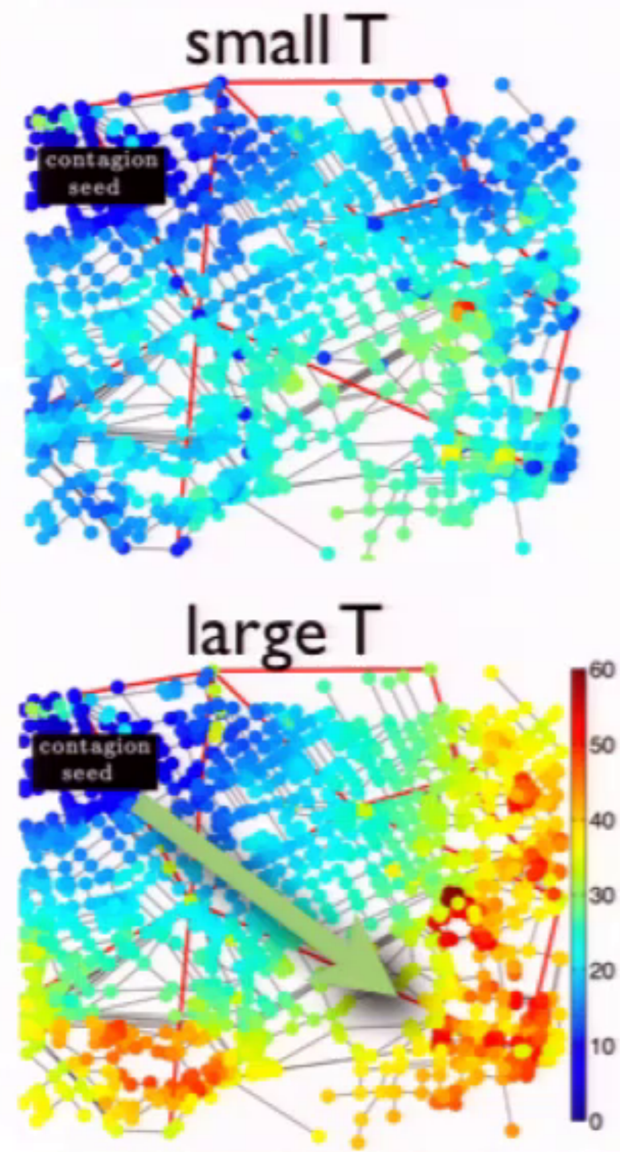
- Performance across the (T, α) parameter space
- Manifold recovered for regime dominated by WFP
- Results shown for $N = 200$, $(d^G, d^{NG}) = (6, 2)$



Applications for a London Transit System

- WFP vs ANC depends on threshold T

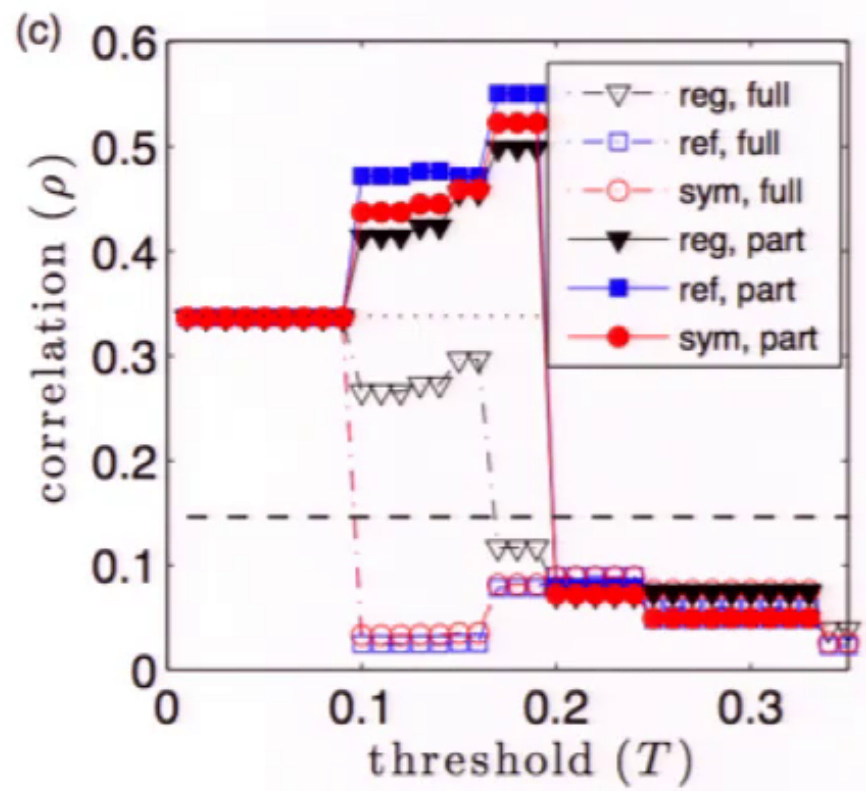




Geometry of WTM Maps

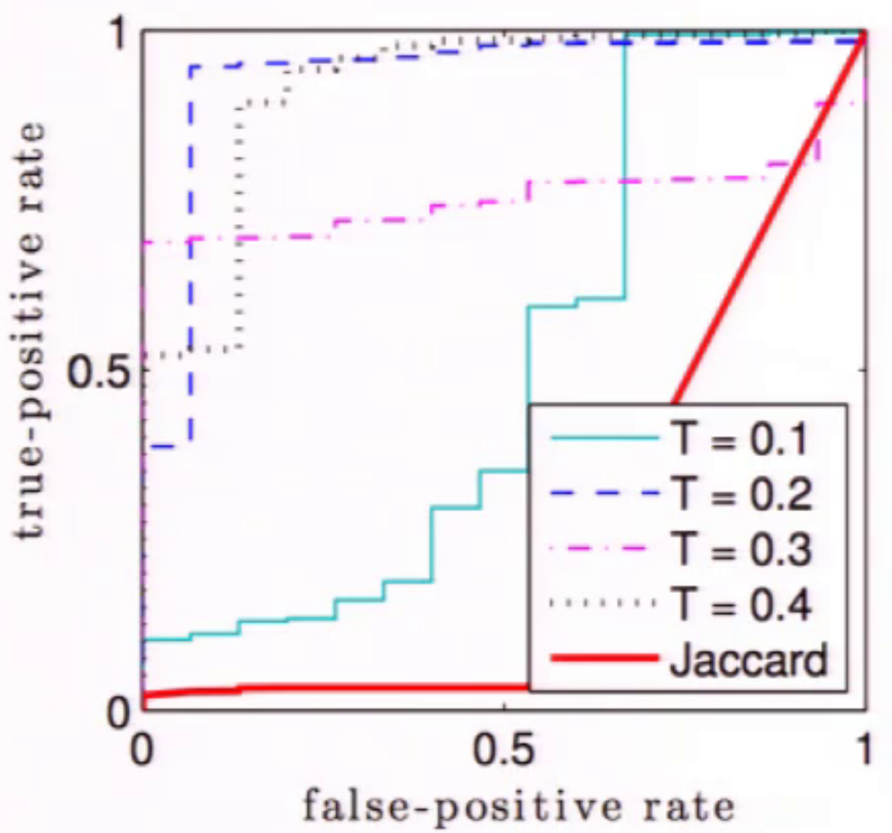
- WTM contagions strongly affected by metro lines
- Contagions follow geometry better for moderate threshold
- Geometry is most distorted for contagions seeded near metro stations

Geometry of WTM-map



For comparison: horizontal lines denote Isomap and Laplacian eigenmap

Denoising Networks with WTM Maps



- Studying geometry leads to a denoising algorithm
- Results shown for the London transit network
- Outperform an approach

Conclusions

- "Contagions Maps" such as WTM maps embed network nodes as a high-dimensional point cloud for analysis
  - Low-dimensional structure in the point cloud can reveal low-dimensional structure in networks
  - Structure in contagion maps also reveals how contagions spread
    - Modeling, forecast, and control of contagions