

Disease spread on networks: integrating structure and dynamics through a generalized inverse

Joseph Tien

Computer Science Seminar, Purdue

September 2017

Joint work

- Pauline van den Driessche
- Zhisheng Shuai
- Marisa Eisenberg

- Karly Jacobsen

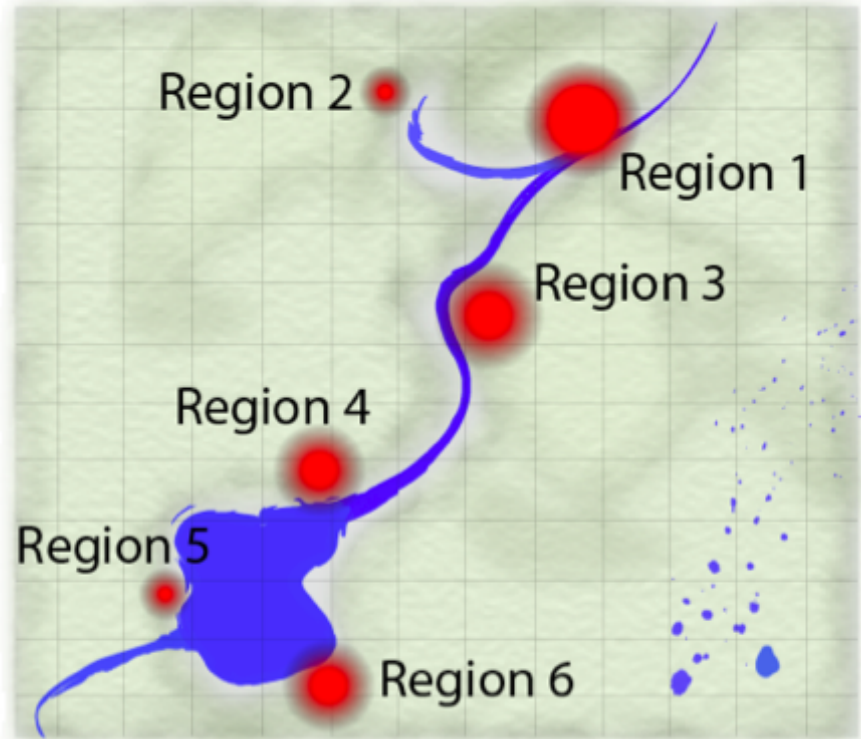
Dynamics on networks

- Large literature on structure of networks
- Also lots of work on how structure informs dynamics on networks
- Less well studied: [how do node dynamics change effective network structure?](#)

Dynamics on networks

- Large literature on structure of networks
- Also lots of work on how structure informs dynamics on networks
- Less well studied: **how do node dynamics change effective network structure?**
- Generalized inverse of the graph Laplacian: **absorption inverse**
 - Absorbing random walks
 - Integrating structure and dynamics
 - Structural metrics: distance, clustering, centrality
 - Implications for disease dynamics on networks

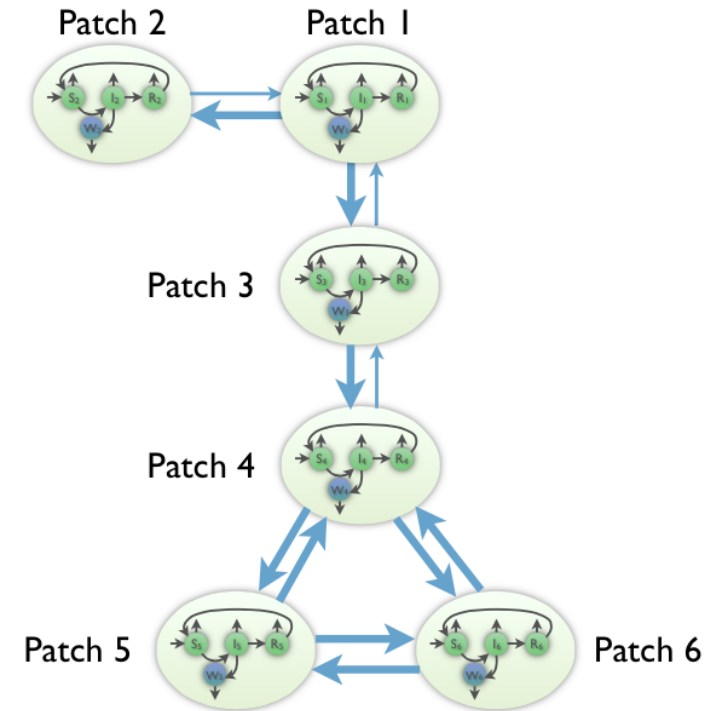
Disease on community networks



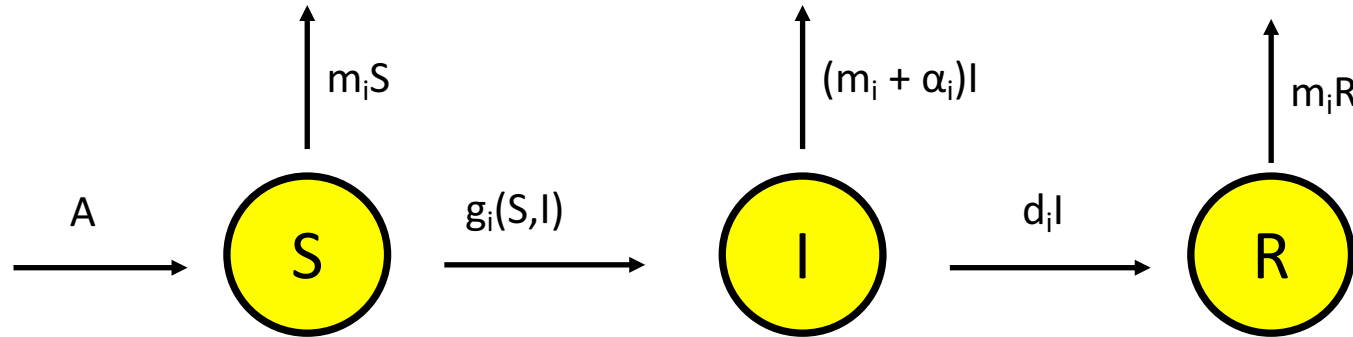
"It has proven difficult to obtain analytical results for metapopulation models." --Riley et al, 2015.

Motivating model framework

- Strongly connected, weighted, directed graph G
- Vertices: low dimensional system of ODEs (“communities”)
- Mobility network (adjacency matrix A)



Vertex dynamics



Pathogen removal rate $d_i + m_i + \alpha_i$ (“absorption”)

$$\mathcal{R}_0^{(i)} = g'(N, 0) \times \frac{1}{d_i + m_i + \alpha_i}$$

Basic reproduction number

R_0 and the next generation matrix

- Basic reproduction number for more complicated settings: next generation matrix
- Diekmann, Heesterbeek, and Metz (1990) – general setting (operator)
- Van den Driessche and Watmough (2002) – compartmental models (matrix)

$$\mathcal{R}_0 = \rho(FV^{-1})$$

F – “fecundity” matrix (new infections)

V – “transfer” matrix

V^{-1} – “fundamental matrix”

Next generation matrix

$$\mathcal{R}_0 = \rho(FV^{-1})$$

F – “fecundity” matrix (new infections)

V – “transfer” matrix

V^{-1} – “fundamental matrix”

V_{ij}^{-1} – expected time in i , starting from j

Theorem (van den Driessche and Watmough):

$\mathcal{R}_0 = 1$ threshold for disease invasion

Next generation matrix

$$\mathcal{R}_0 = \rho(FV^{-1}) \quad \text{Hard to analyze!}$$

F – “fecundity” matrix (new infections)

V – “transfer” matrix

V^{-1} – “fundamental matrix”

V_{ij}^{-1} – expected time in i , starting from j

Theorem (van den Driessche and Watmough):

$R_0 = 1$ threshold for disease invasion

Domain R_0 for community networks

$$\begin{aligned}\dot{I} &= D_S D_\beta I - (L + D)I, \\ V &= L + D\end{aligned}$$

L – graph Laplacian

$$L = W - A$$

Laplacians and graph structure

- Un-normalized Laplacian (L)
- Normalized Laplacian ($W^{-1/2}LW^{-1/2}$)
- Random walk Laplacian ($W^{-1}L$)
- Encoding of structural information!
 - Number of connected components
 - Spectral gap, Cheeger's inequality
 - Matrix tree theorem
 - Community detection, centrality, more...

Matrix Tree Theorem

Let (\mathcal{G}, A) be a weighted, directed graph, and let L be the Laplacian matrix of (\mathcal{G}, A) . Let c_{kk} denote the (k, k) cofactor of L . Then the cofactors of L are related to the rooted spanning trees of \mathcal{G} by the following:

$$c_{kk} = \sum_{\mathcal{T} \in \mathbb{T}_k} \prod_{(j,i) \in E(\mathcal{T})} a_{ij}, \quad (1)$$

where \mathbb{T}_k is the set of all spanning in-trees rooted at vertex k , $E(\mathcal{T})$ is the arc set of rooted spanning in-tree \mathcal{T} , and a_{ij} the weight of the arc from j to i .

Matrix Tree Theorem

Let (\mathcal{G}, A) be a weighted, directed graph, and let L be the Laplacian matrix of (\mathcal{G}, A) . Let c_{kk} denote the (k, k) cofactor of L . Then the cofactors of L are related to the rooted spanning trees of \mathcal{G} by the following:

$$c_{kk} = \sum_{\mathcal{T} \in \mathbb{T}_k} \prod_{(j,i) \in E(\mathcal{T})} a_{ij}, \quad (1)$$

where \mathbb{T}_k is the set of all spanning in-trees rooted at vertex k , $E(\mathcal{T})$ is the arc set of rooted spanning in-tree \mathcal{T} , and a_{ij} the weight of the arc from j to i .

Corollary: $\ker L$ can be expressed in terms of spanning trees of G

$$\ker L = u$$

$u_i =$ normalized sum of weights of spanning trees rooted at i

Laurent series expansion for R_0

Fundamental matrix as perturbation of Laplacian:

$$V = L + \varepsilon D$$

Scaling:

Time scales of absorption to movement

Langenhop (1971): Laurent series for perturbed singular matrices

$$V^{-1}(\varepsilon) = \frac{1}{\varepsilon}X_{-1} + X_0 + \varepsilon X_1 + \varepsilon^2 X_2 + \dots$$

X_{-1} – spanning trees, averaging

X_0 – generalization of the group inverse; higher order structure

Singular term X_{-1} : averaging on the network

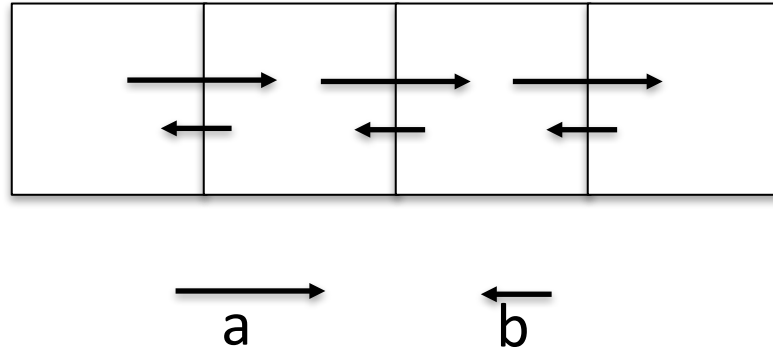
$$\begin{aligned}\mathcal{R}_0 &\approx \frac{1}{\bar{d}} \sum_{i=1}^n \mathcal{R}_0^{(i)} u_i d_i \\ &= \mathbb{E}[\mathcal{R}_0^{(i)}]\end{aligned}$$

u_i – weight of spanning trees rooted at i

d_i – absorption rate at i

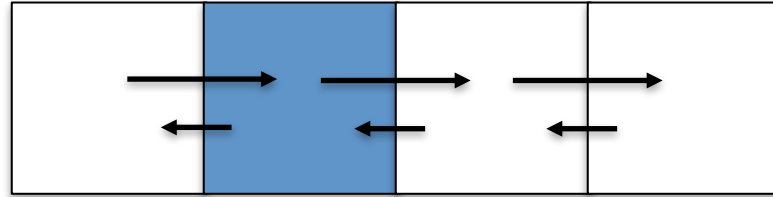
$\frac{u_i d_i}{\bar{d}}$ – probability measure combining structure and dynamics

Rooted spanning trees: “rivers”



u – PageRank vector, no teleportation

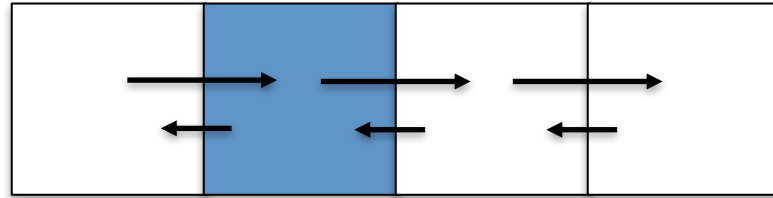
Rooted spanning trees: “rivers”



\xrightarrow{a}

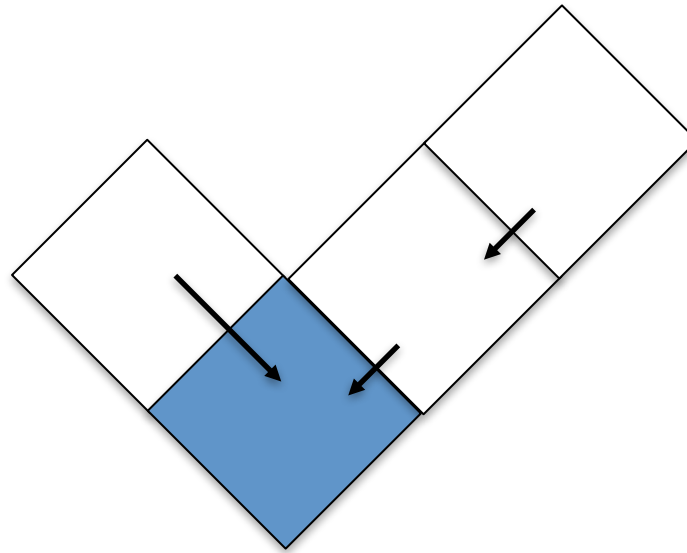
\xleftarrow{b}

Rooted spanning trees: “rivers”



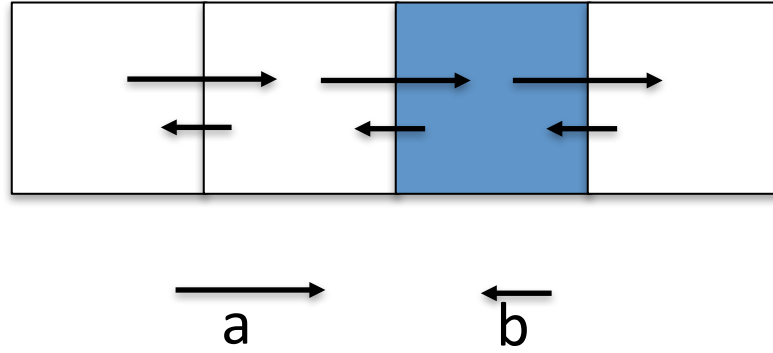
\xrightarrow{a}

\xleftarrow{b}

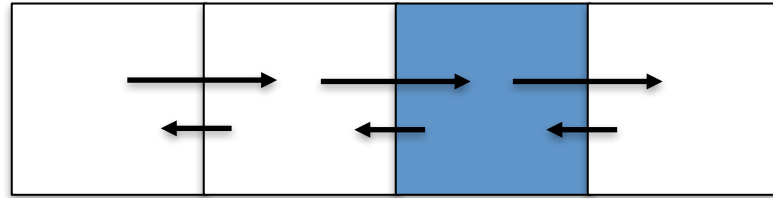


Tree weight:
 ab^2

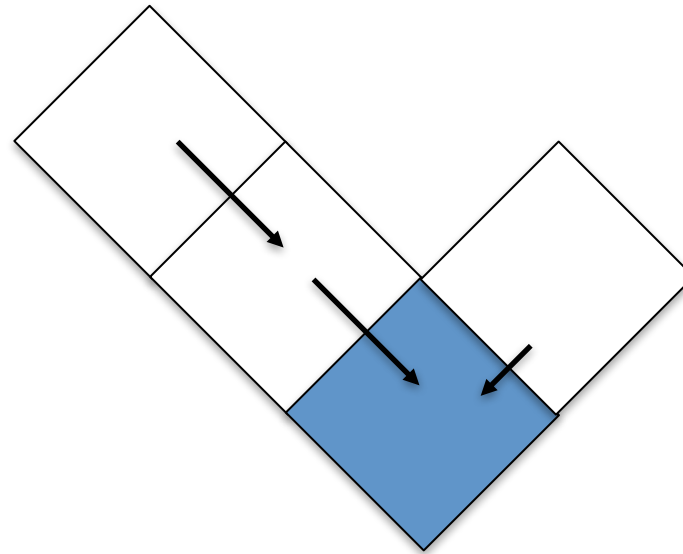
Rooted spanning trees: “rivers”



Rooted spanning trees: “rivers”

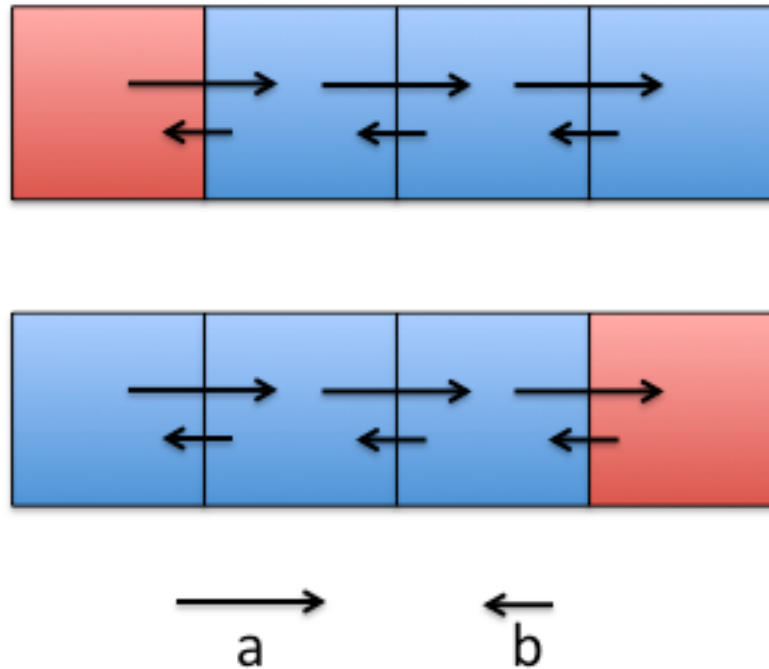


\xrightarrow{a} \xleftarrow{b}



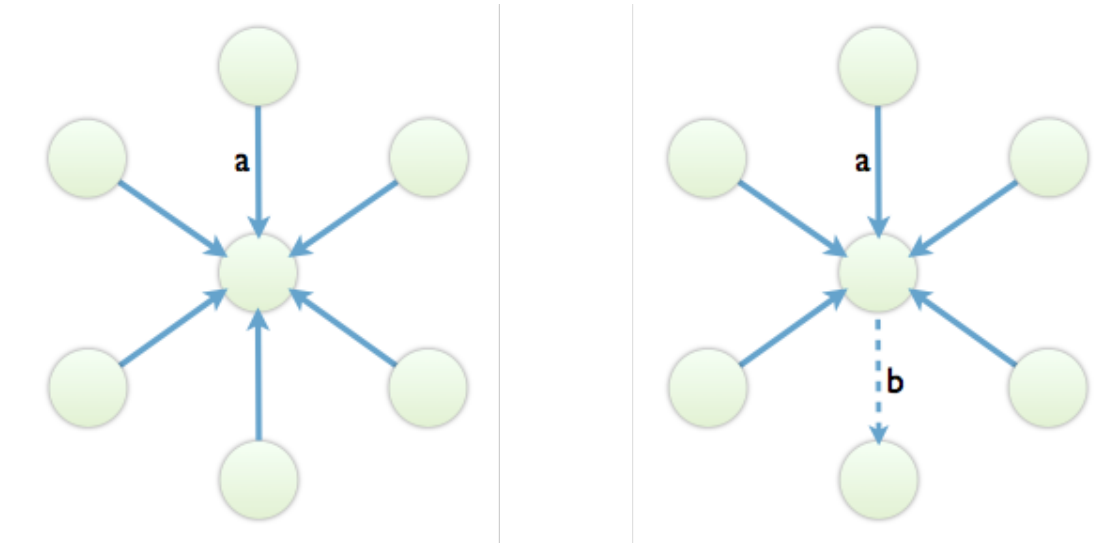
Tree weight:
 a^2b

Network motif: “rivers”



- Network risk increases by a factor of a/b each step downstream
- Worst place for disease hot spot -- **downstream**

Network motif: Star



$$u_{hub} = \frac{a}{a + nb}$$

Balanced graphs

- **Balanced graph** – the net outflow equals net inflow for every vertex
 - Generalization of symmetric network
- **Identical network risk** for every vertex:
 $u_i = u_j$ for all i, j

Network risk: net inflow vs. net outflow

High net inflow vs outflow = high network risk

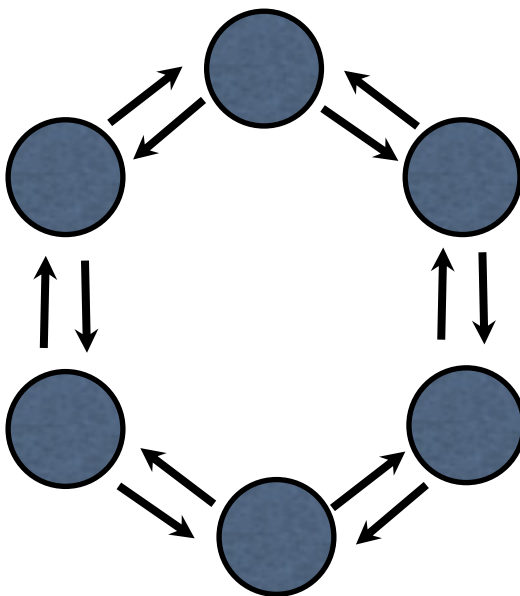
Beyond averaging

- Analytical results involving lowest order approximation are clean, provide biological insights
- Balanced graphs – uniform u
- Information outside the radius of convergence? $\varepsilon \gg 1$
- Higher order network structure?

Beyond averaging

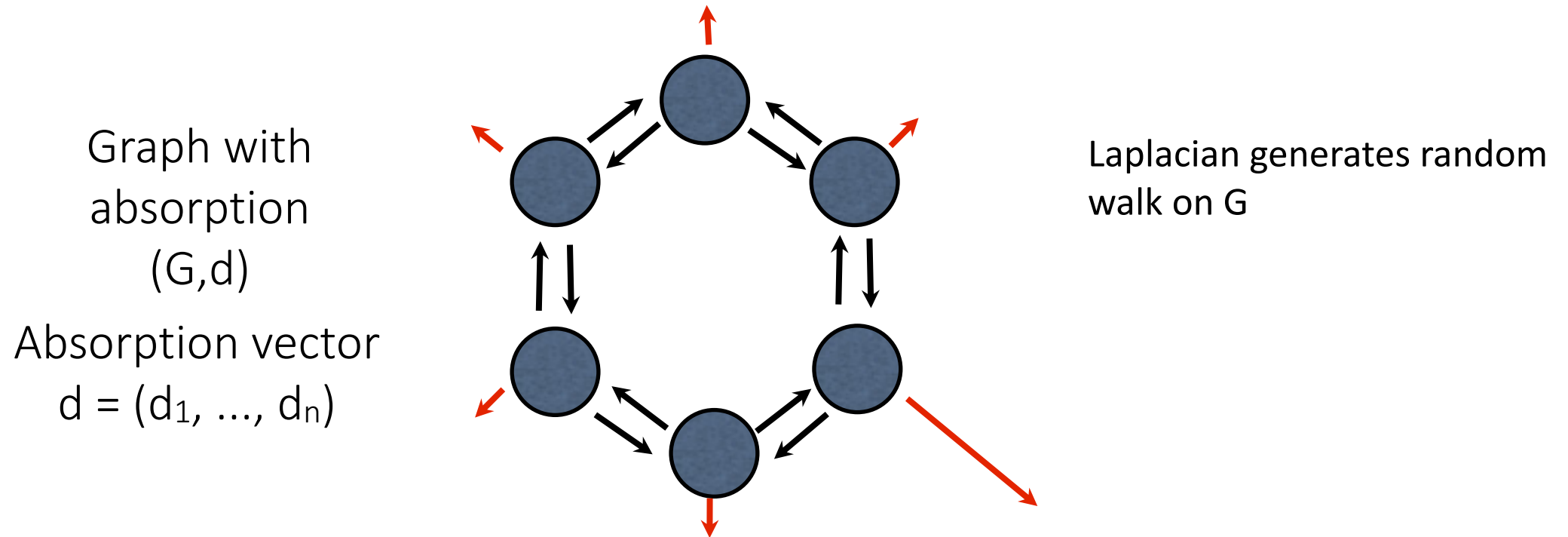
- Analytical results involving lowest order approximation are clean, provide biological insights
- Balanced graphs – uniform u
- Information outside the radius of convergence? $\varepsilon \gg 1$
- Higher order network structure?
- X_0 – absorption inverse
 - Basic connections for [graphs with absorption](#)
 - Metrics based upon the absorption inverse combining graph structure and dynamics at vertices

Graphs with absorption and absorbing random walks



Laplacian generates random
walk on G

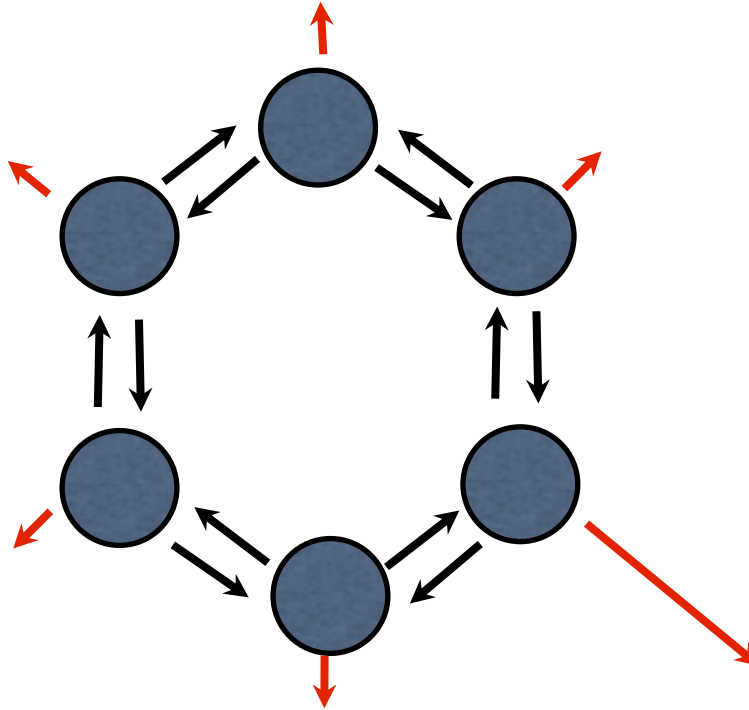
Graphs with absorption and absorbing random walks



Graphs with absorption and absorbing random walks

Graph with
absorption
 (G, d)

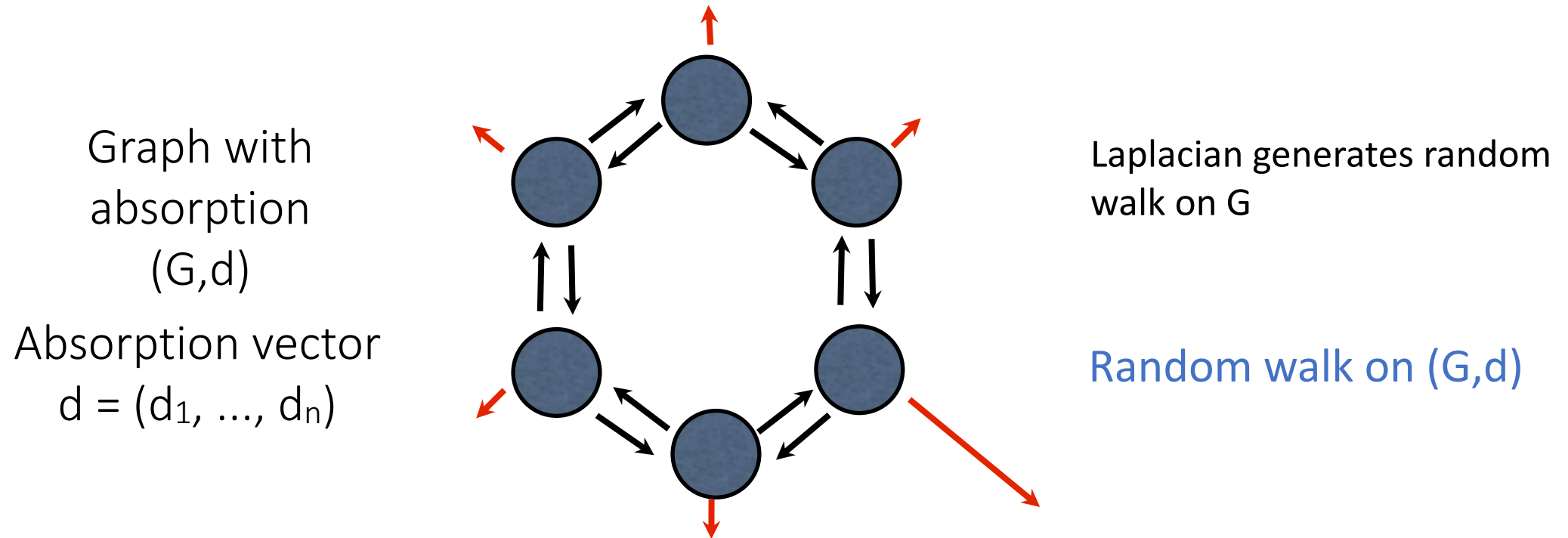
Absorption vector
 $d = (d_1, \dots, d_n)$



Laplacian generates random
walk on G

Random walk on (G, d)

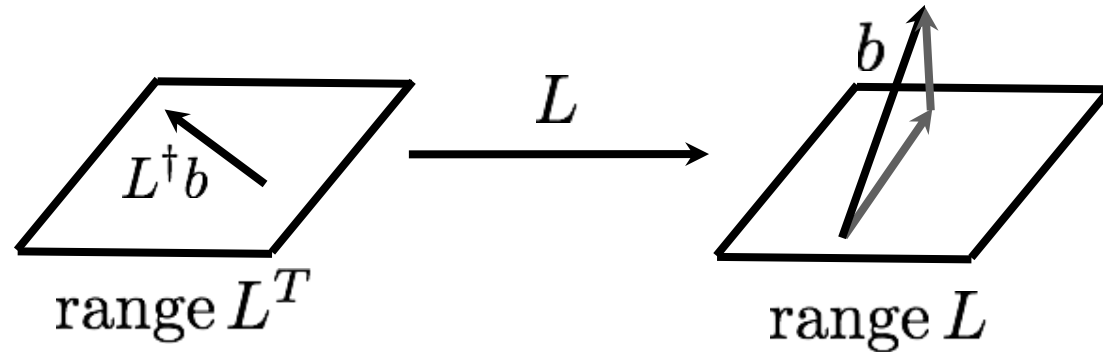
Graphs with absorption and absorbing random walks



Absorption inverse – (new) generalized inverse for (G, d)

Generalized inverses

- Different generalized inverses for different problems
- Least squares solutions -- Moore-Penrose inverse



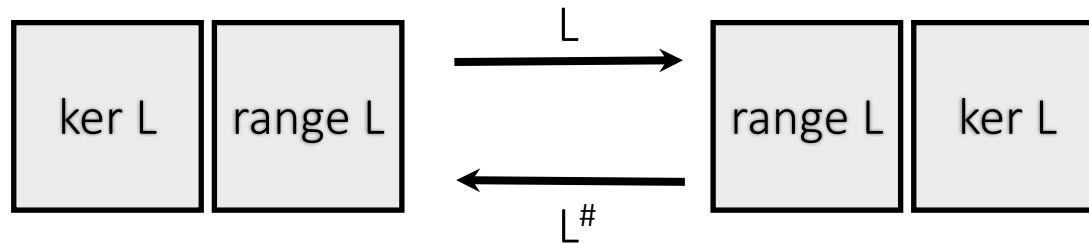
$$Lx = b,$$

$$b \notin \text{range } L$$

L^\dagger – Moore-Penrose inverse of L

Generalized inverses

- Group inverse $L^\#$ -- inverts on range L
- Square matrix L of index one
- Spectral inverse of L



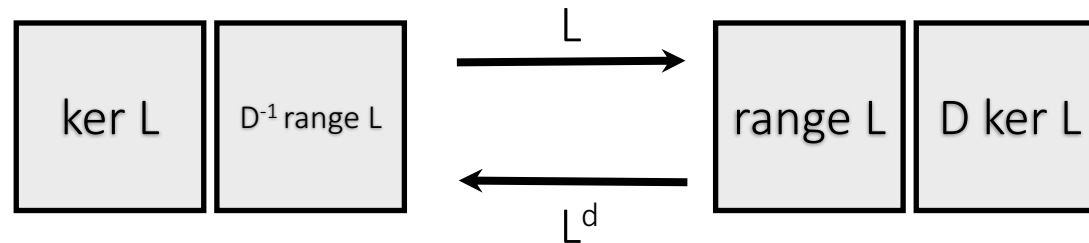
$$\mathbb{R}^n = \ker L \oplus \text{range } L$$

The absorption inverse

Definition. Let L be the Laplacian of a strongly connected graph with absorption. Then the absorption inverse L^d is the matrix which satisfies

$$L^d L y = y \text{ for } y \in D^{-1}(\text{range } L)$$

$$L^d y = 0 \text{ for } y \in D(\ker L)$$



L^d exists, is unique, and is equal to X_0

Generalized inverses and graph structure

- Laplacian L generates random walk on G
- Commute distance and L^\dagger (Lovasz; Klein and Randic; Boley et al)
- Spectral clustering and $L^\#$ (Fiedler; Hagen and Kahn)
- PageRank and $L^\#$ (Chung)

Properties of the absorption inverse

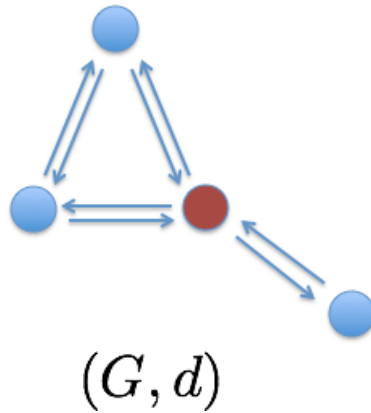
- $L^d = L^\#$ iff absorption rates are equal
- $L^d = L^\dagger$ if the absorption rates are equal and G is balanced
- $L^d + (L^d)^T$ positive semidefinite for balanced graphs

Theorem. The absorption inverse L^d can be expressed in terms of any $\{1\}$ -inverse Y of L :

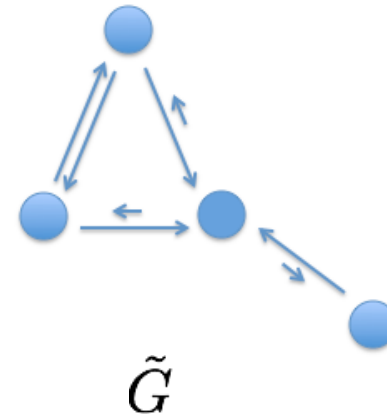
$$L^d = \left(I - \frac{1}{\hat{d}} U D \right) Y \left(I - \frac{1}{\hat{d}} D U \right)$$

L^d and the absorption-scaled graph

G – original graph



A – adjacency matrix $\tilde{A} = AD^{-1}$

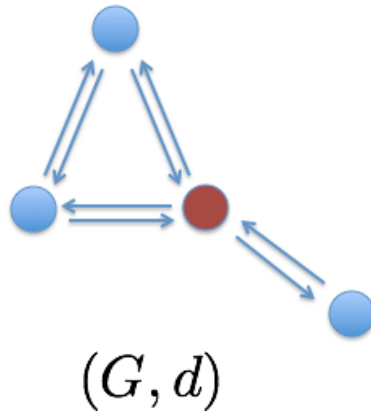


\tilde{G} – absorption-scaled graph

$$L^d = D^{-1} \tilde{L}^\#$$

L^d and the absorption-scaled graph

G – original graph

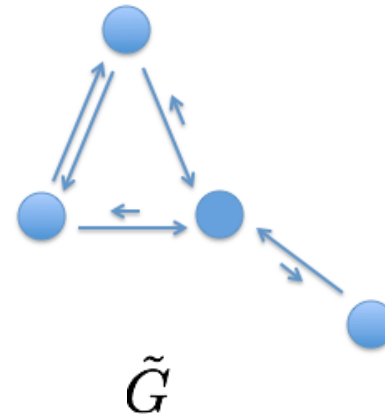


(G, d)

A – adjacency matrix

$$\tilde{A} = AD^{-1}$$

\tilde{G} – absorption-scaled graph



\tilde{G}

$$L^d = D^{-1} \tilde{L}^\#$$

Probability measure = $\ker L^d$

L^d and the absorption-scaled graph

$$\lim_{z \rightarrow 0} (L + zI)^{-1} L = L^\# L \quad (\text{Ben-Israel})$$

$$\lim_{z \rightarrow 0} (L + zD)^{-1} L = L^d L.$$

L^d and the absorption-scaled graph

Resolvent and
group inverse

$$\lim_{z \rightarrow 0} (L + zI)^{-1} L = L^\# L \quad (\text{Ben-Israel})$$

$$\lim_{z \rightarrow 0} (L + zD)^{-1} L = L^d L.$$

$$\begin{aligned} (L + zD)^{-1} &= D^{-1} (LD^{-1} + zI)^{-1} \\ &= D^{-1} (\tilde{L} + zI)^{-1}. \end{aligned}$$

$$D^{-1} \tilde{L}^\# = L^d$$

What good is it?

- **Mathematics:** combinatorial matrix theory and L^d
- **Network science:** directed distance metric
- **Applications:** clustering and contagion

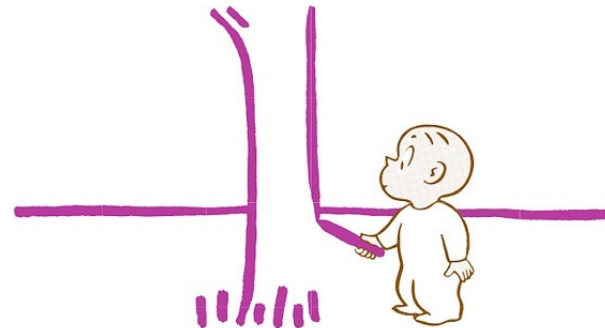
Forests and the Laplacian

Definition: A **forest** is a collection of trees.

Matrix forest theorems: Chaiken (1982)

Forests and eigenvalues of Laplacian of subgraphs: Chung (2010)

Spanning forests and the group inverse: Kirkland et al (1997), Chebotarev and Agaev (2002)



A forest consisting of a single tree

Forests and the group inverse

Theorem 1 (Chebotarev and Agaev) *For any $\tau \in \mathbb{R}$,*

$$(I + \tau L)^{-1} = \frac{1}{\sigma(\tau)} (Q_0 + \tau Q_1 + \dots + \tau^{n-1} Q_{n-1})$$

where $\sigma(\tau) = \sum_{k=0}^{n-1} \sigma_k \tau^k$.

Q_k — matrix of in-forests with k arcs

$$[Q_k]_{ij} = \omega(\mathcal{F}_k^{j \rightarrow i})$$

σ_k — the weight of all in-forests with k arcs

A forest theorem for L^d

Theorem 1 *Let L^d be the absorption inverse for the graph with absorption (G, d) and let \tilde{G} be the corresponding absorption-scaled graph. Then,*

$$L_{ij}^d = \frac{\omega(\tilde{\mathcal{F}}_{n-2}^{j \rightarrow i})}{d_i \tilde{\sigma}_{n-1}} - \frac{\tilde{\sigma}_{n-2} u_i}{\tilde{\sigma}_{n-1} \bar{d}}. \quad (1)$$

A forest theorem for L^d

Theorem 1 *Let L^d be the absorption inverse for the graph with absorption (G, d) and let \tilde{G} be the corresponding absorption-scaled graph. Then,*

$$L_{ij}^d = \frac{\omega(\tilde{\mathcal{F}}_{n-2}^{j \rightarrow i})}{d_i \tilde{\sigma}_{n-1}} - \frac{\tilde{\sigma}_{n-2} u_i}{\tilde{\sigma}_{n-1} \bar{d}}. \quad (1)$$

Follows from Chebotarev and Agaev and the relationship between L^d and the absorption-scaled graph

A forest theorem for L^d

Theorem 1 *Let L^d be the absorption inverse for the graph with absorption (G, d) and let \tilde{G} be the corresponding absorption-scaled graph. Then,*

$$L_{ij}^d = \frac{\omega(\tilde{\mathcal{F}}_{n-2}^{j \rightarrow i})}{d_i \tilde{\sigma}_{n-1}} - \frac{\tilde{\sigma}_{n-2} u_i}{\tilde{\sigma}_{n-1} \bar{d}}. \quad (1)$$

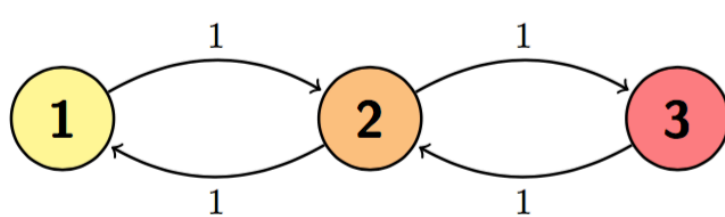
Follows from Chebotarev and Agaev and the relationship between L^d and the absorption-scaled graph

Interpretation: closeness! (Graphs without decay -- Chebotarev and Shamis, 1998)

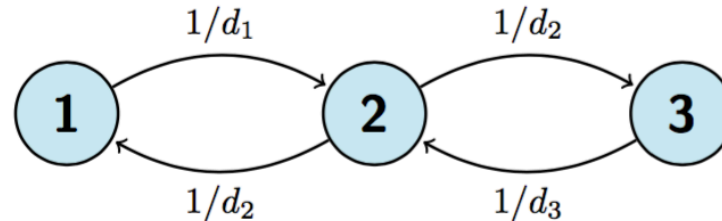
L_{ij}^d large (positive) $\Rightarrow (i, j)$ close

L_{ij}^d small (negative) $\Rightarrow (i, j)$ far

Example: counting spanning forests



(a)



(b)

$\Gamma \in \tilde{\mathcal{F}}_{n-1}$	$\omega(\Gamma)$	$\Gamma \in \tilde{\mathcal{F}}_{n-2}$	$\omega(\Gamma)$
	$\frac{1}{d_1 d_2}$		$\frac{1}{d_1}$
	$\frac{1}{d_2 d_3}$		$\frac{1}{d_2}$
	$\frac{1}{d_1 d_3}$		$\frac{1}{d_2}$
			$\frac{1}{d_3}$

Distance metric for (G,d)

Large L^d \Rightarrow Small distance
Small (negative) L^d \Rightarrow Large distance

$$R(j, i) = \begin{cases} K - [L^d]_{ij} & i \neq j \\ 0 & i = j \end{cases}$$

$$K = \max_i L_{ii}^d$$

Directed distance: distance from j to i

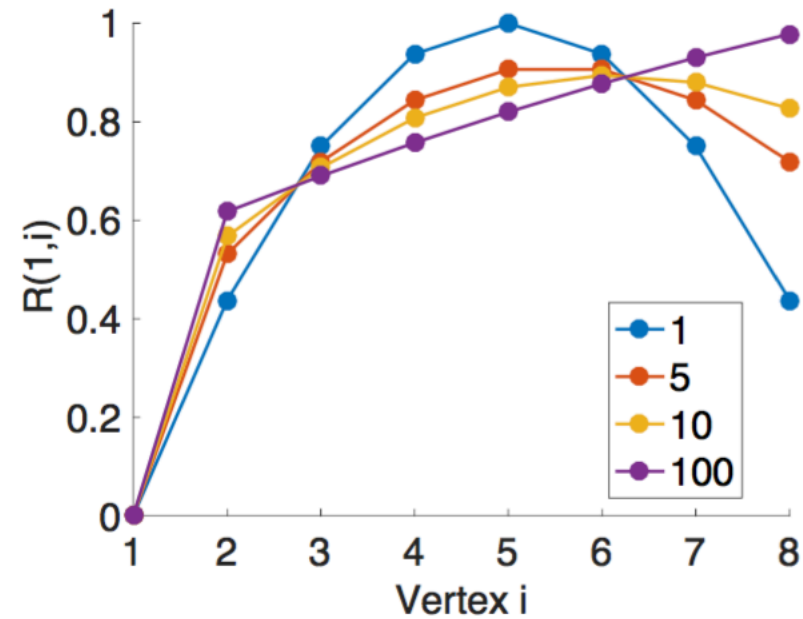
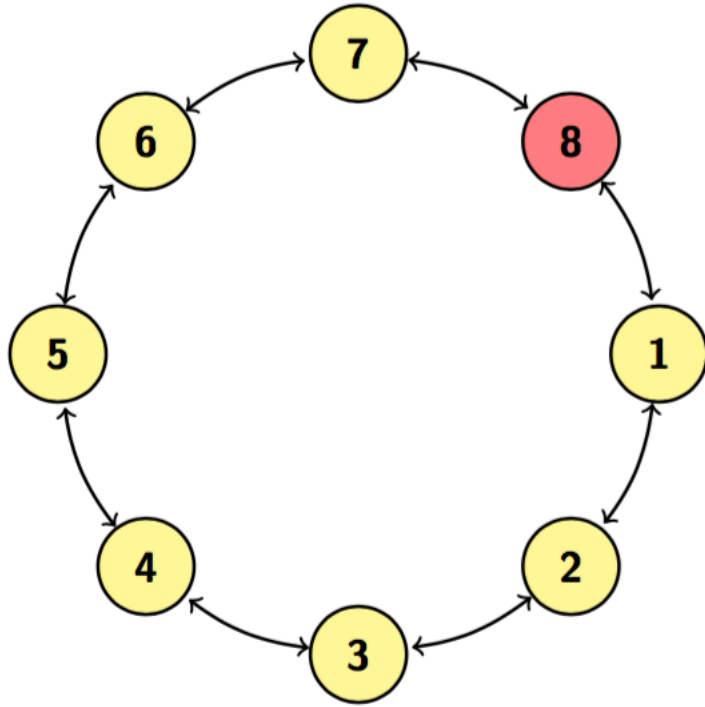
Distance metric for (G,d)

- Triangle inequality depends upon diagonal dominance of L^d

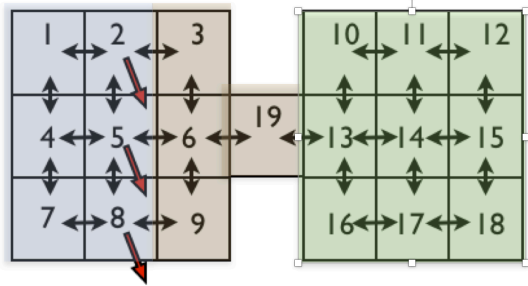
Theorem 1 *L^d has the property of (row) diagonal maximality. That is, for each i , $L_{ij}^d < L_{ii}^d$ for all $j \neq i$.*

- Distance metric reflects node dynamics (absorption), unlike existing metrics

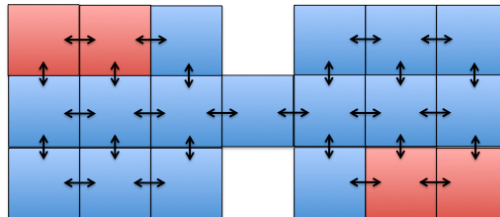
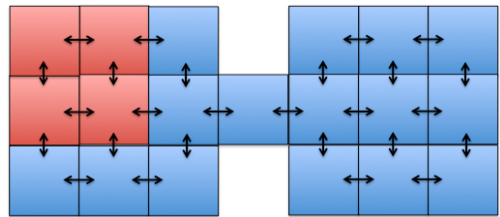
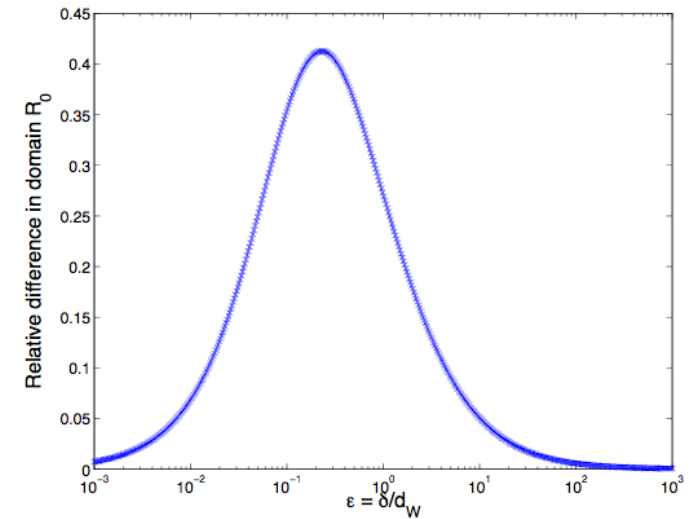
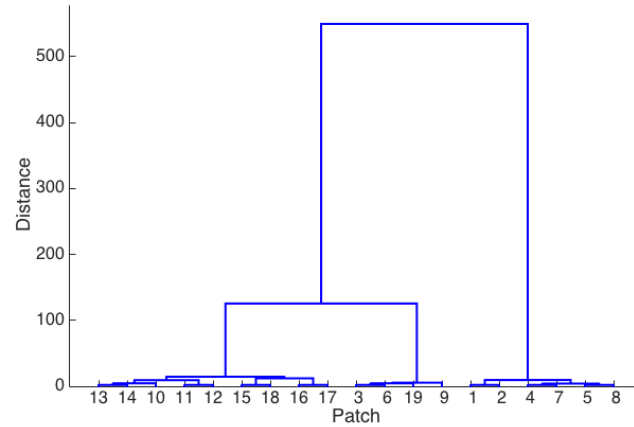
Example: absorption can effectively change topology



Hierarchical clustering, communities, and R_0



K. Jacobsen



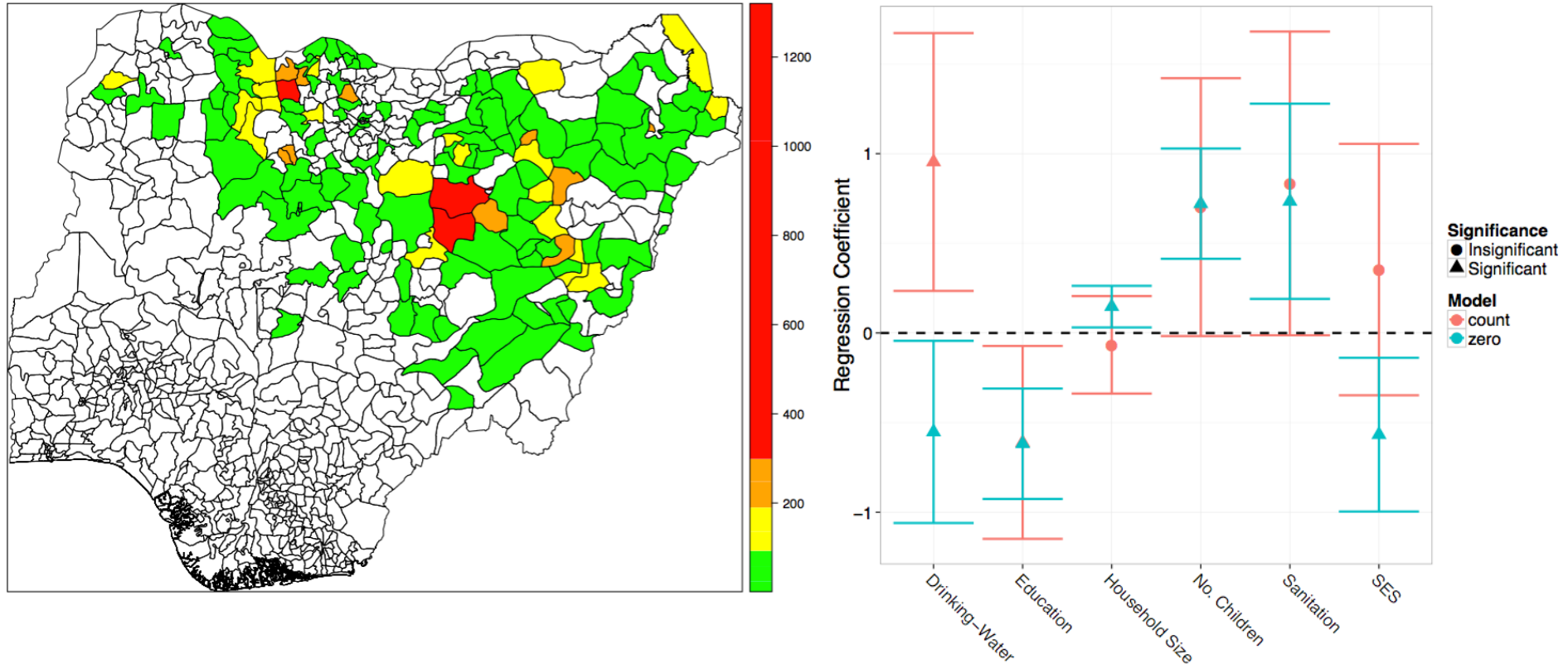
$$\Delta \mathcal{R}^{(i)} = (\text{patch risk of } i) \times E_i[L^d]$$

$$\mathcal{R}_0 \approx \hat{\mathcal{R}}_0 + \sum_{i=1}^n \Delta \mathcal{R}^{(i)}$$

Application to empirical networks

- Data on network structure
- Data on node characteristics (transmissibility, absorption)

Case study: cholera in Nigeria



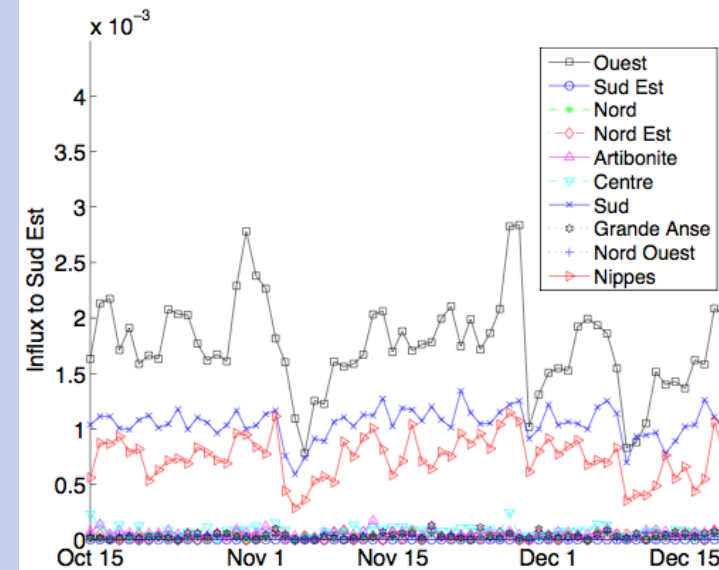
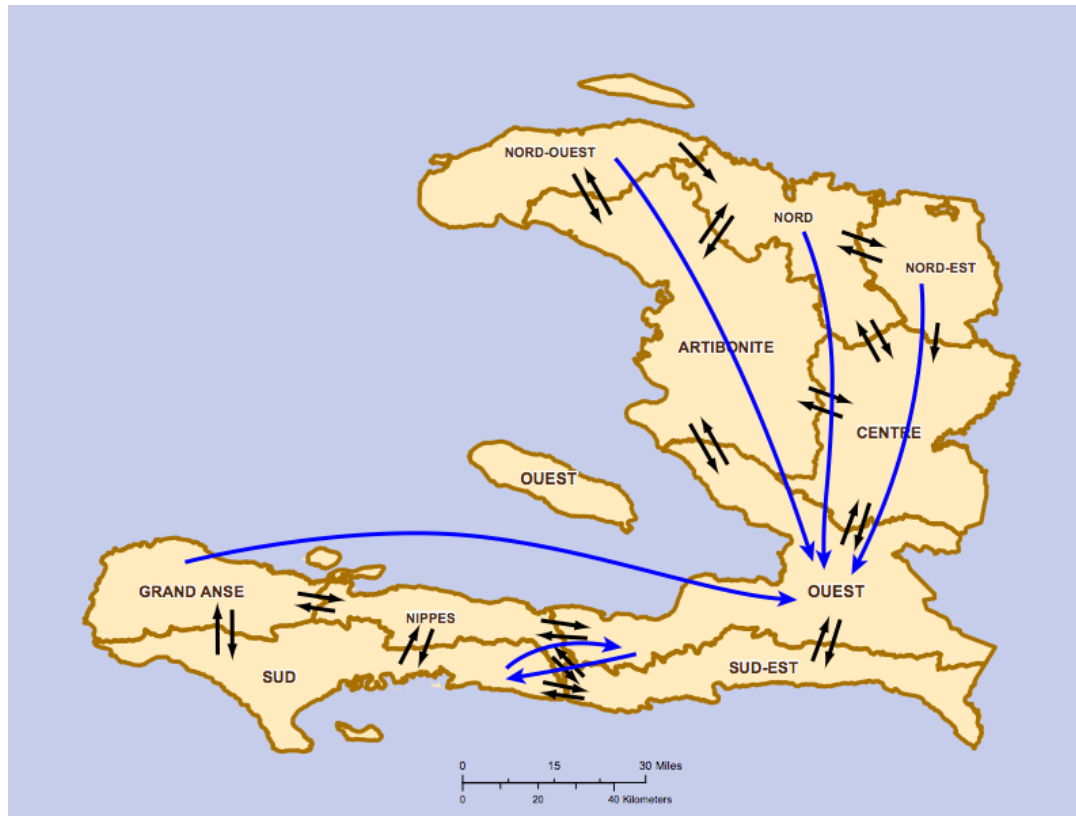
Regression of **attack rate** vs. **DHS covariates**:
Strong relationship with **education**
Weak / unclear relationship with sanitation, drinking water

Potential for using basic demographic covariates to estimate patch transmissibility



Estimating mobility networks

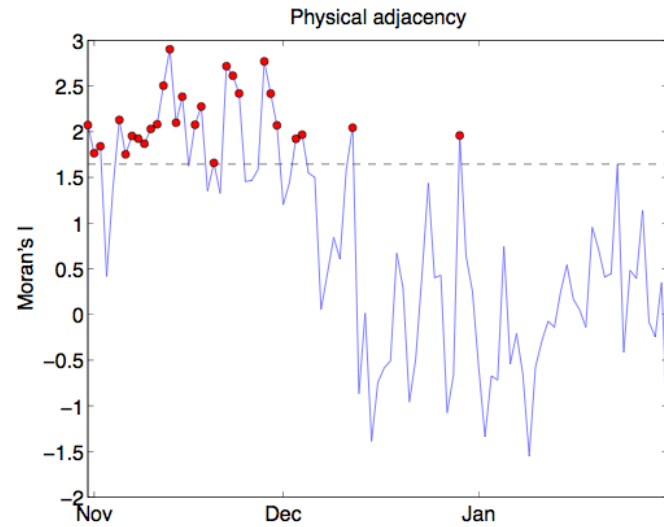
- Many rich data sets exist regarding network structure and mobility....



Cell phone movements,
Haiti Oct 15-Dec 15, 2010
Flowminder (L. Bengtsson et al)

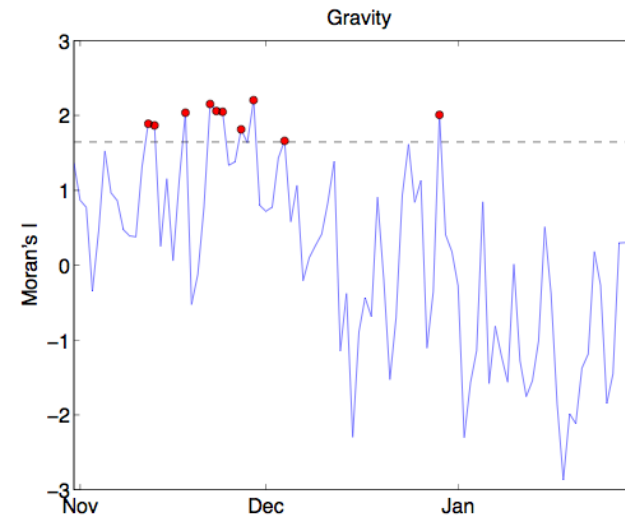
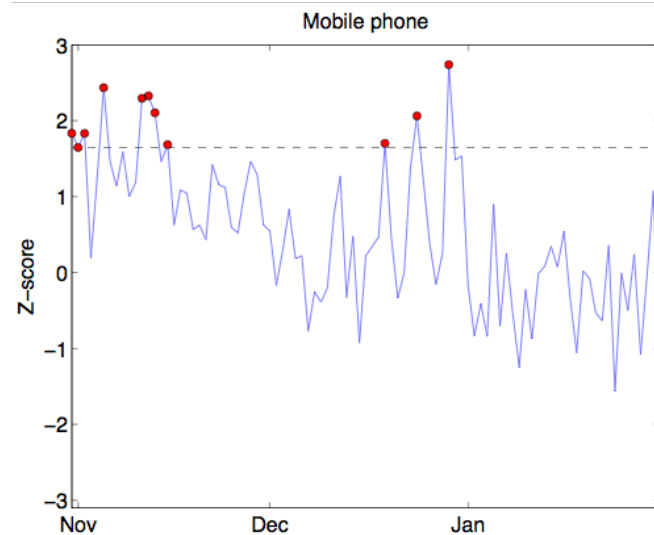
Estimating mobility networks

-but getting data on how *infected* individuals move is a challenge!

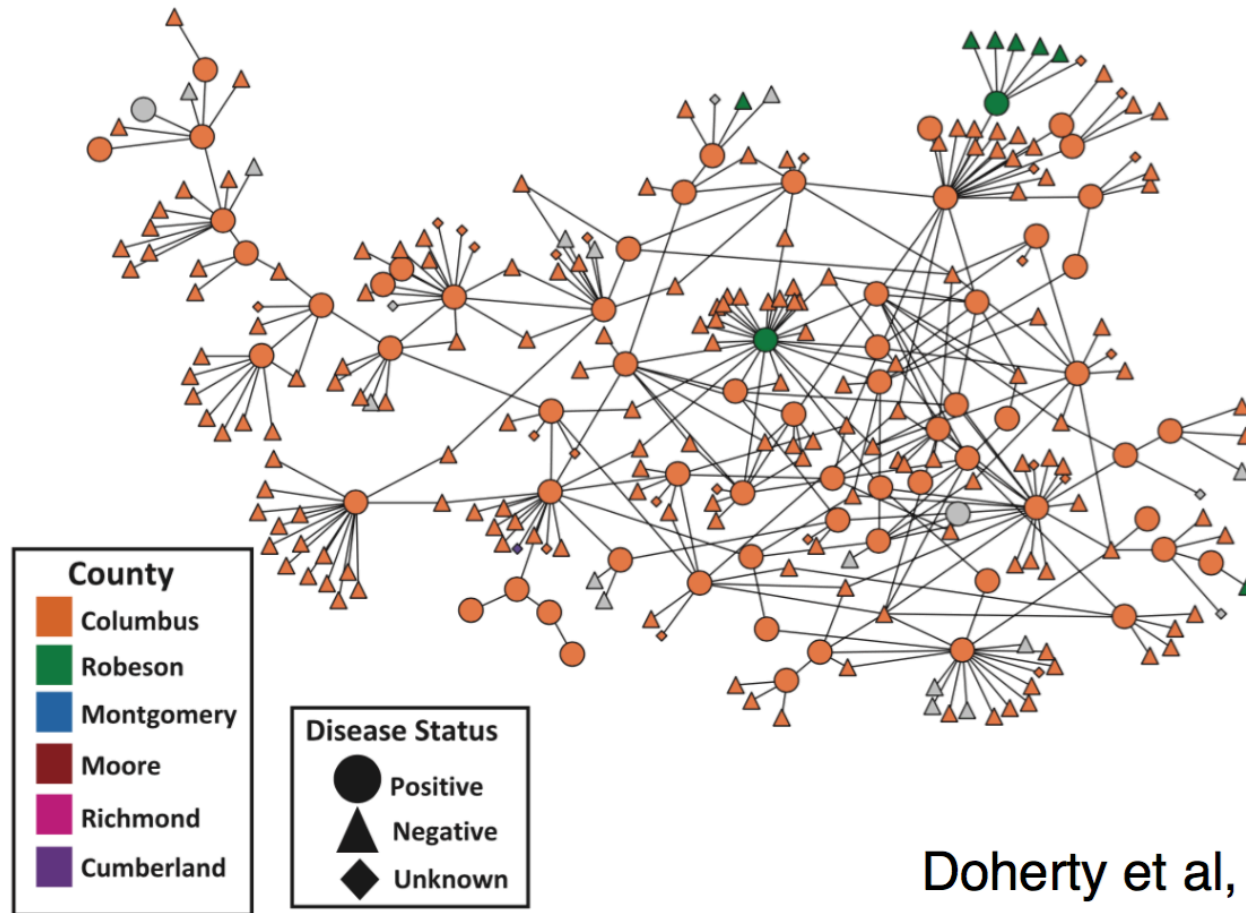


Moran's I spatial correlation

Weights according to different networks



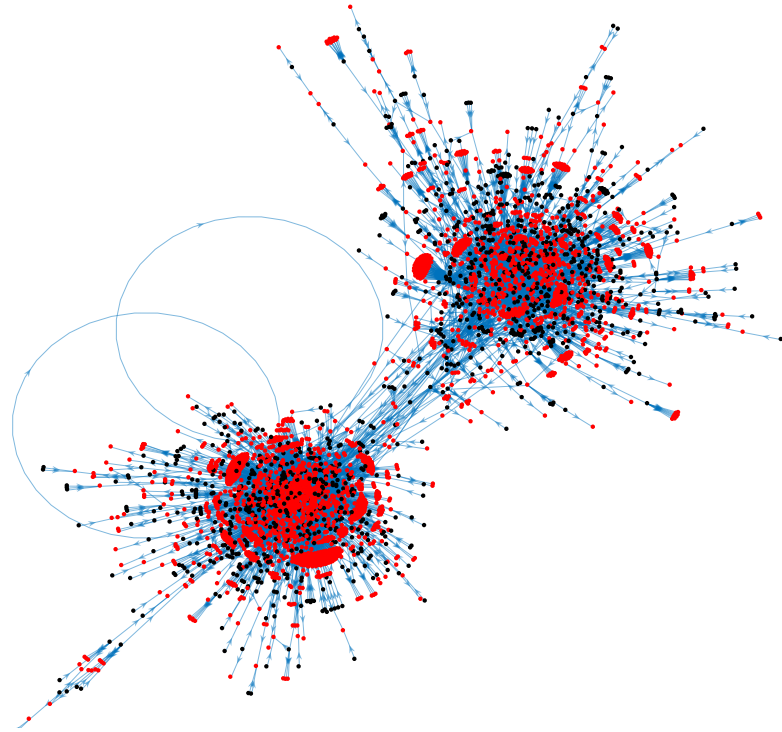
Sexual networks and core groups



Collaborators:
Bill Miller,
Abby Norris Turner

Doherty et al, 2012. Epidemiology

Social media networks and information propagation



Twitter re-tweet network, #Charlottesville

Acknowledgements

Karly Jacobsen

Leah Durbak

Linus Bengtsson

Zhisheng Shuai

WHO Cholera Task Force

Marisa Eisenberg

Nigeria Ministry of Health

Pauline van den Driessche

National Science Foundation
Mathematical Biosciences Institute
Centers for Disease Control

References

Jacobsen K, Tien JH. A generalized inverse for graphs with absorption. arXiv: 1611.02233.

Tien JH, Shuai Z, Eisenberg MC, van den Driessche P. 2015. J. Math. Biology 70: 1065-1092.

Durbak L, Oyemakinde A, Tien JH. Demographic covariates of cholera risk in Nigeria, 2010. In revision.