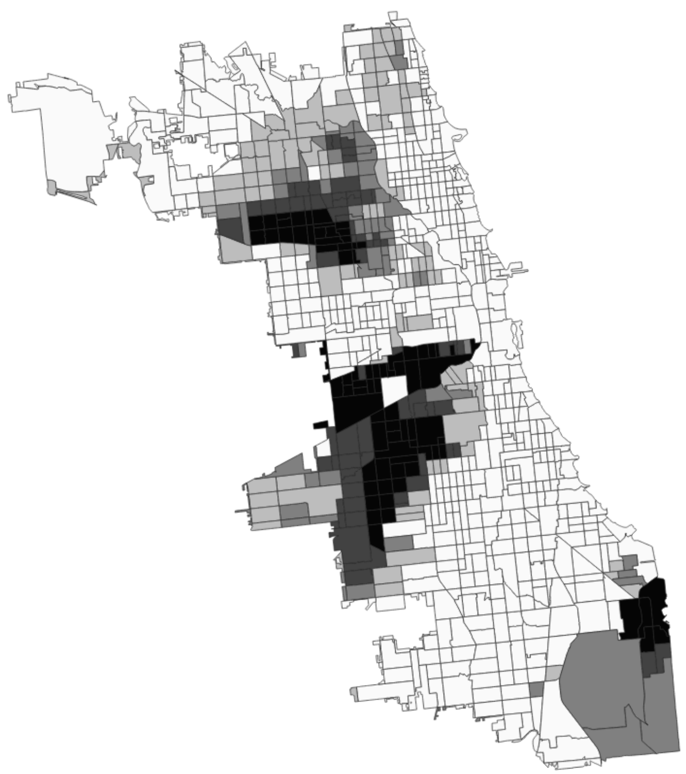
# Analytic Methods for Measuring Network Segregation

*James M. Murphy*
Department of Mathematics
**October 6, 2022**

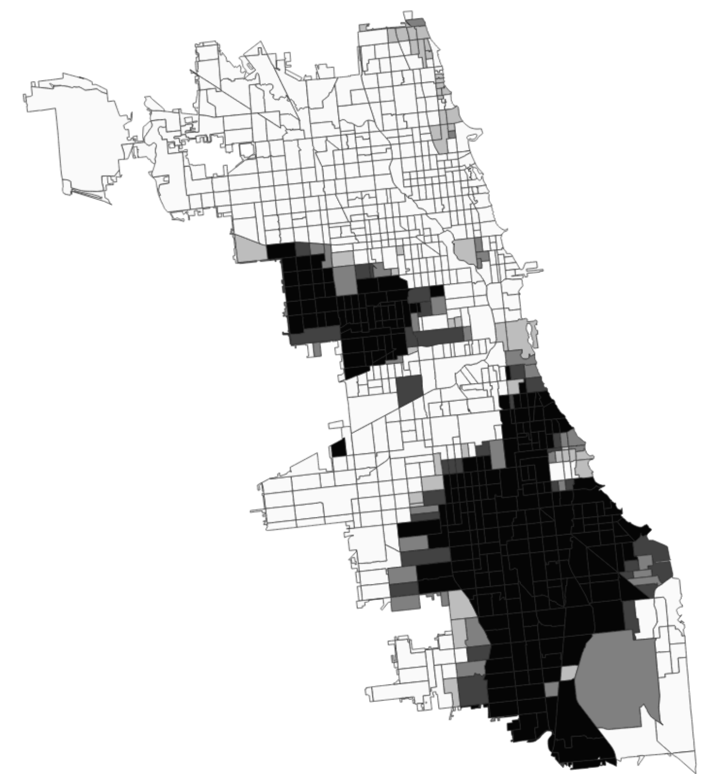
Tufts University

# Measures of Segregation

- How to think about segregation on graphs coming from geography?



Hispanic Population in Chicago



Black Population in Chicago

- What metrics capture the intuition that these populations are highly segregated?

# From Intuitive Description to Formal Math

- "A geographic area is segregated when most of a unit's neighbors are in the same group as the unit."
- Fine, but vague. What is a region? Who are neighbors? What is a group?
- Our approach is to use graph theory and functions on graphs to formalize this. Then, we can start asking mathematical questions and proposing new approaches.

Tufts UNIVERSITY

# What Should a Segregation Metric Do?

1. Capture intuition in important cases.

   *"This population in this city looks segregated and my metric agrees"*

2. Allow for comparisons on the same geographic region.

   *"In City 1, population A is more segregated than population B"*

3. Allow for comparisons across different geographic regions.

   *"Population A is more segregated in City 1 than in City 2"*

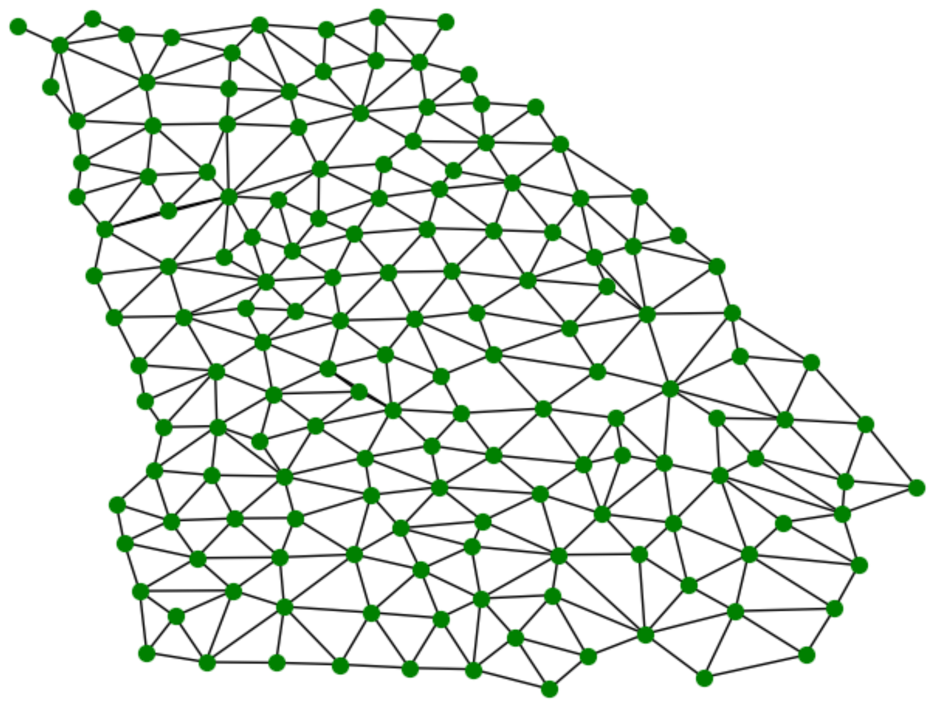
4. Admit theoretical guarantees on performance.

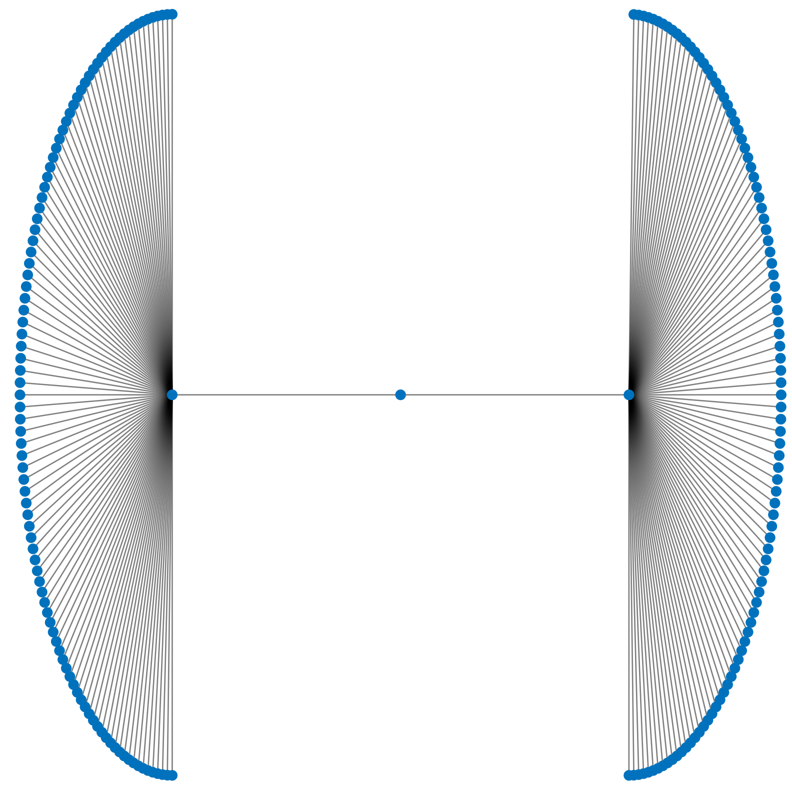
*"Any population with X properties will be classified as segregated by my metric."*

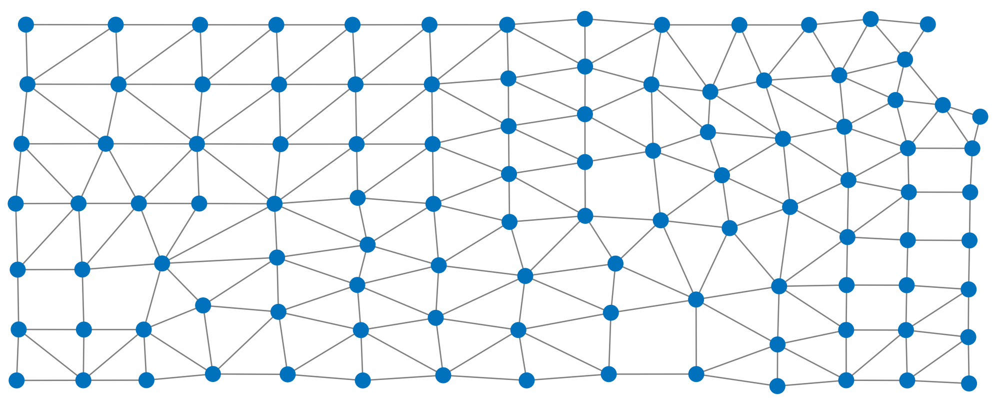
Tufts UNIVERSITY

# Informal Definition

A graph $\mathcal{G}$ is a collection of nodes and edges between nodes.



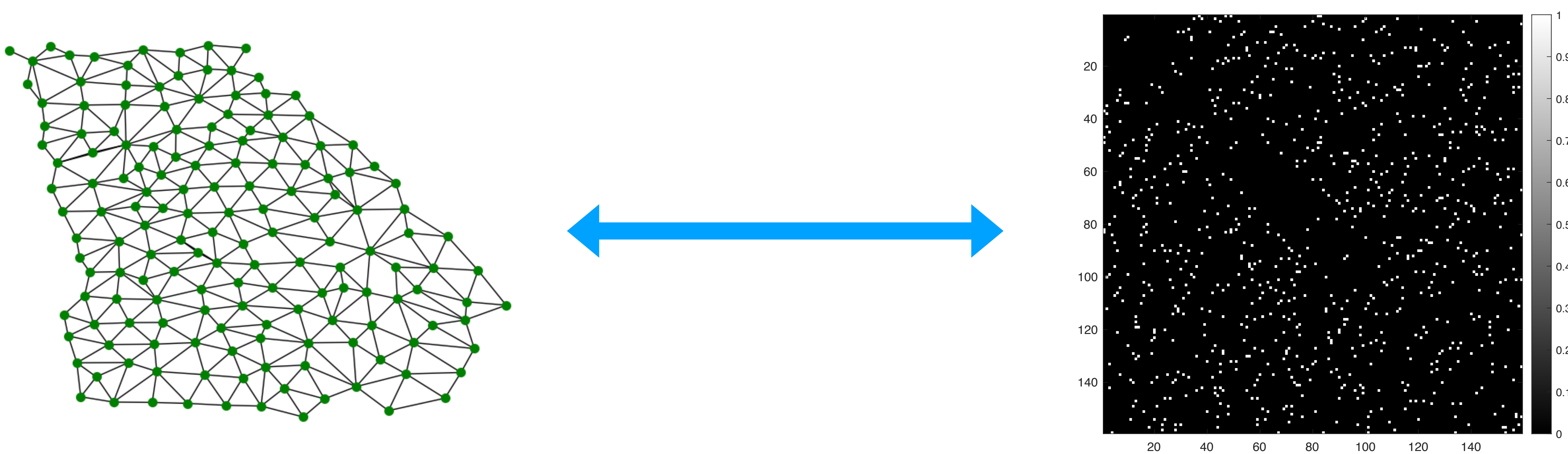




Tufts UNIVERSITY

# Adjacency Matrix

- A graph: nodes (e.g., census tracts) with edges between them (e.g., an edge between adjacent tracts that touch).
- Let $\mathcal{G}$ have $n$ nodes. Let $A$ be an $n \times n$ matrix with $A_{ij} = 0$ if there is no edge between the $i^{th}$ and $j^{th}$ nodes and $A_{ij} = 1$ if there is.




Tufts UNIVERSITY

# Moran's $\mathtt{I}$

- Our jumping-off point is *Moran's* $\mathtt{I}$, which takes in a graph and function on the graph and gives a number that can be interpreted as measuring segregation.

**Definition.** *Let* $W \in \mathbb{R}^{n\times n}$ *be a matrix that is not the zero matrix, and let* $w = \sum_{i,j=1}^{n} |W_{ij}|$. *Let* $\bar{\mathsf{v}}$ *denote the mean of a vector* $\mathsf{v}$. Moran's $\mathtt{I}$ with respect to $W$ *is a functional* $\mathtt{I}(\,\cdot\,;W) : \mathbb{R}^n \to \mathbb{R}$ *defined by*

$$\mathtt{I}(\mathsf{v};W) := \left(n\sum_{i,j=1}^{n} W_{ij}(v_i-\bar{\mathsf{v}})(v_j-\bar{\mathsf{v}})\right) \Big/ \left(w\sum_{i=1}^{n}(v_i-\bar{\mathsf{v}})^2\right).$$

- Most commonly, we take $W = A$. Other choices have interesting properties as well.

# Intuition and Toy Example

# Does Moran's $\mathtt{I}$ Work?

- The major claims in the geography literature around $\mathtt{I}$ are that it takes values in $[-1, 1]$ with

$$\mathtt{I}(\mathsf{v}; A) \approx 1 \quad \longleftrightarrow \quad \mathsf{v} \text{ is highly segregated}$$

- We wanted to *prove* these results.

- Main finding: they are roughly true when the graph (encoded by $A$) is highly structured, but not when $A$ is irregular.

# Spectrum of Graph Determines $\mathtt{I}$ Range

**Theorem.** *Let $A$ be the adjacency matrix of an undirected graph $\mathcal{G}$. Then the range of possible $\mathtt{I}$ values satisfies* $\mathtt{I}(X; A) \subseteq \left[\frac{\lambda_n}{\bar{\mathsf{d}}}, \frac{\lambda_1}{\bar{\mathsf{d}}}\right]$.

- Here, $\lambda_n, \lambda_1$ are the smallest and largest eigenvalues of the adjacency matrix.
- Here, $\bar{\mathsf{d}}$ is the average degree (number of edges at a node) of the graph.
- So, this result establishes the conventional wisdom pertaining to Moran's I in the case when $-\lambda_n = \bar{\mathsf{d}} = \lambda_1$. This is exactly when the graph is *regular*.

# Regular Graphs

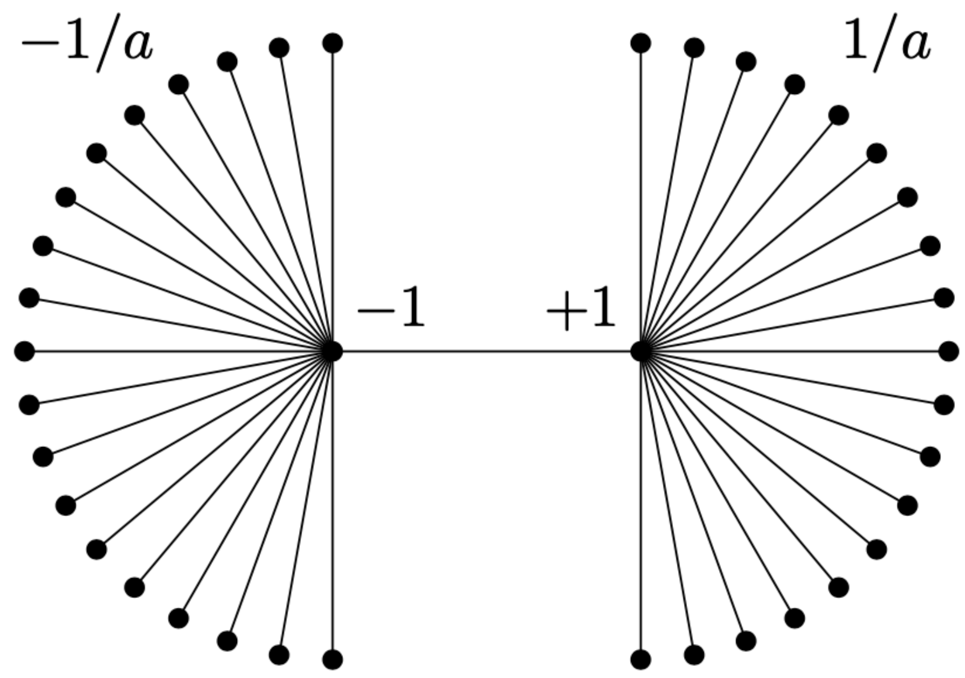
- If all nodes in a graph have the same number of edges, we say the graph is *regular*.

**Corollary.** *Let $A$ be the adjacency matrix of an undirected, regular graph $\mathcal{G}$. Then the range of possible* $\mathtt{I}$ *values satisfies* $\mathtt{I}(X; A) \subseteq [-1, 1]$.

- This follows from bounding the eigenvalues of the graph in terms of the degree when the graph is regular.
- So, the folklore result on how to understand $\mathtt{I}$ is true at least when the graph is regular.

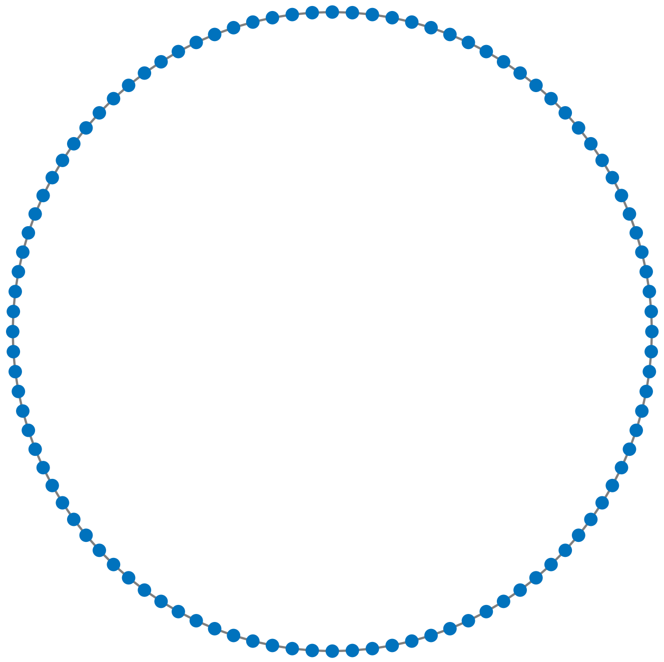
Tufts UNIVERSITY

# Contrived Counterexample

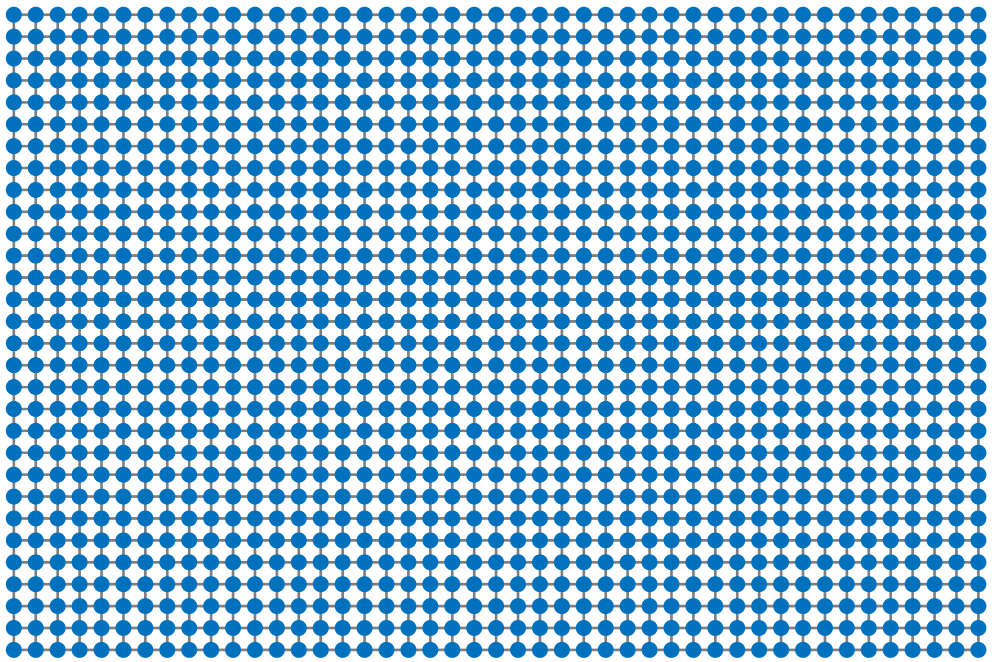



- Looks nothing like a graph coming from geography.
- But, when the number of nodes is large, `I` approaches $a$.
- In particular, `I` can be made arbitrarily large or small!
- The problem is the very high degree nodes—extremely irregular.

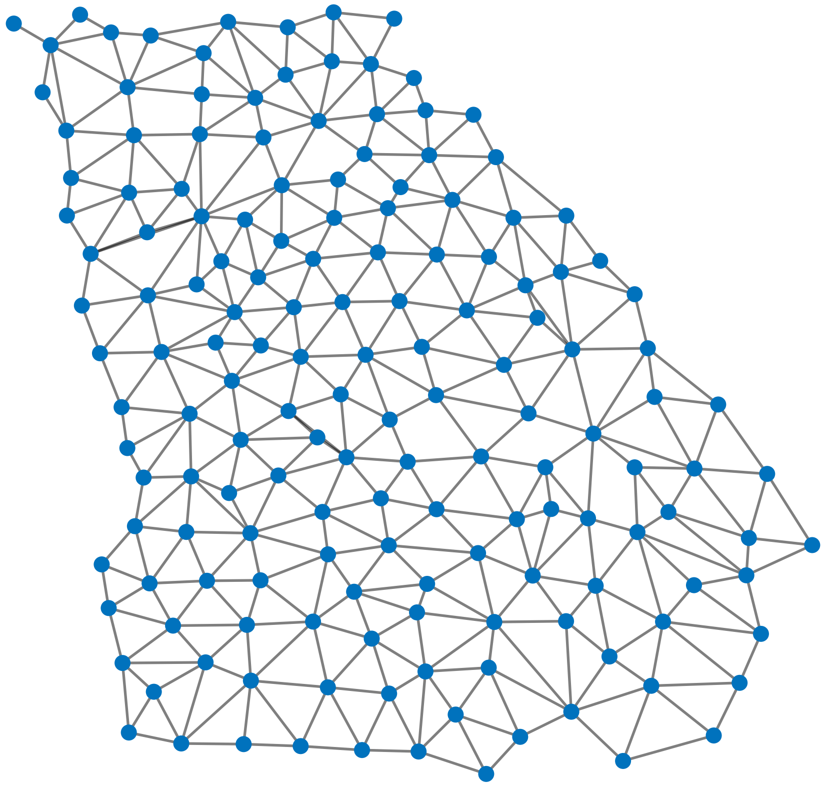
# Are Geography Graphs Close to Regular?



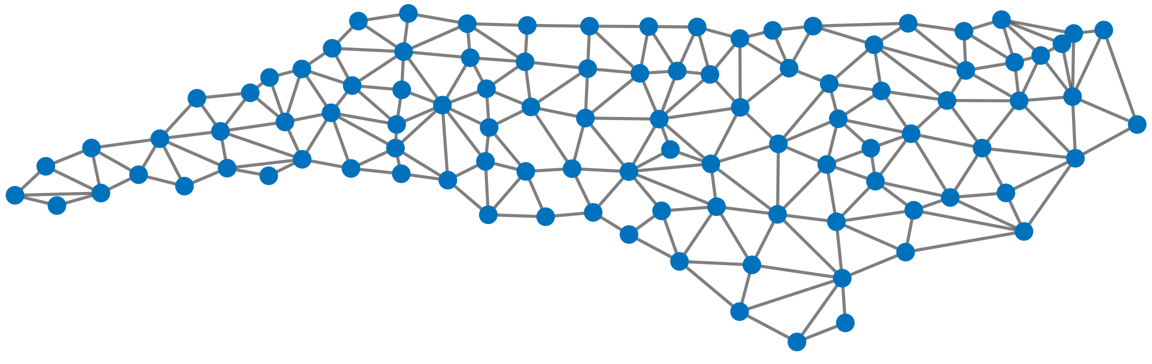
2-Regular



Almost 4-Regular



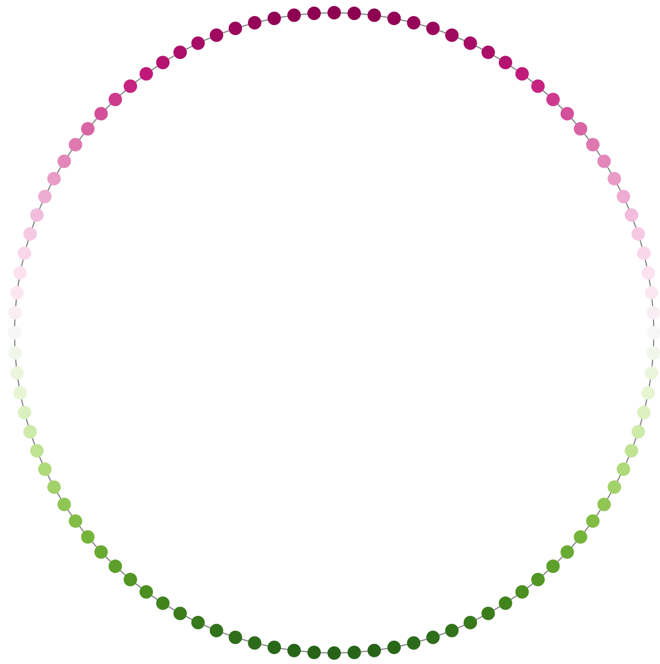
Not So Regular…



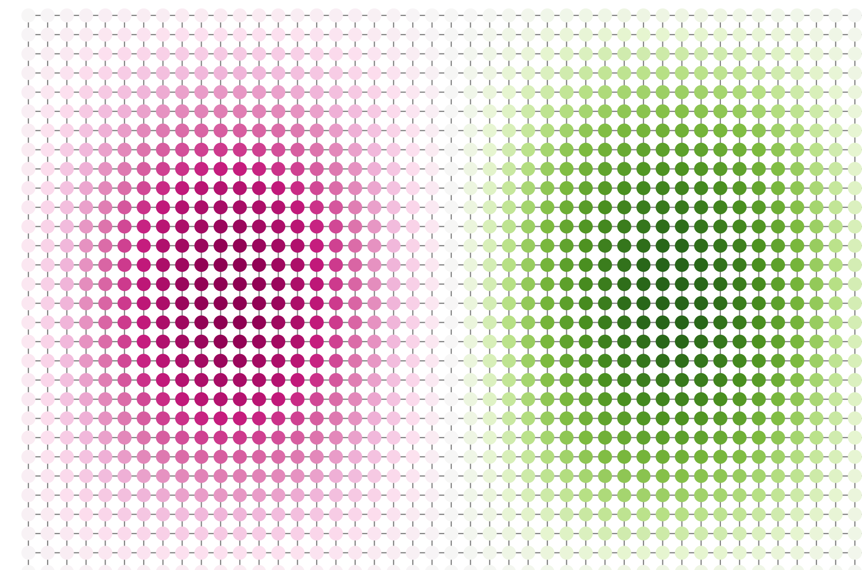
Not So Regular…

# Actual Maximizers (Computed Numerically)

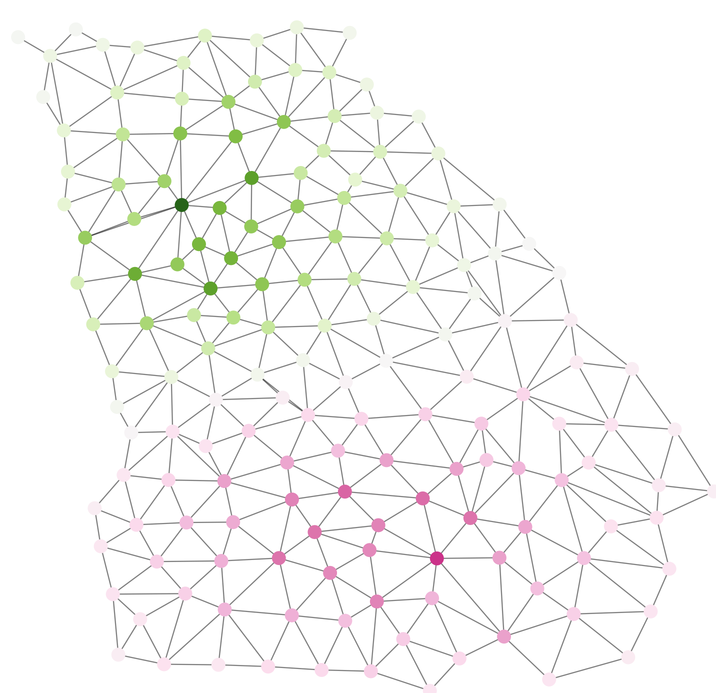
$v = v^A_{max},\ I(v;A) = 0.99803$



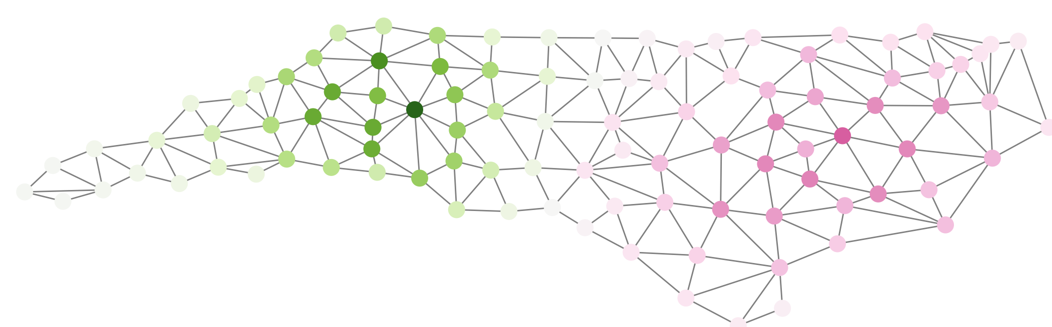
$v = v^A_{max},\ I(v;A) = 1.0211$



$v = v^A_{max},\ I(v;A) = 1.0763$



$v = v^A_{max},\ I(v;A) = 1.1034$



# Generalized Eigenvector Formulation

**Theorem.** *Let $W$ be a symmetric $n \times n$ weight matrix and let $\Pi$ be the projection onto the space of mean 0 vectors. Let $\{(\lambda_i, \Phi_i)\}_{i=1}^{n-1}$ be $\Pi$-orthonormal generalized eigenvectors for the pair $(\Pi W \Pi, \Pi)$. Then for all non-zero $\mathsf{v} \in \mathbb{R}^{n\times 1}$,*

*(a)* $\mathsf{v} = \left(\sum_{i=1}^{n-1} \alpha_i \Pi \Phi_i\right) + \bar{v}\mathbb{1}$, *for some coefficients* $\{\alpha_i\}_{i=1}^{n-1}$.

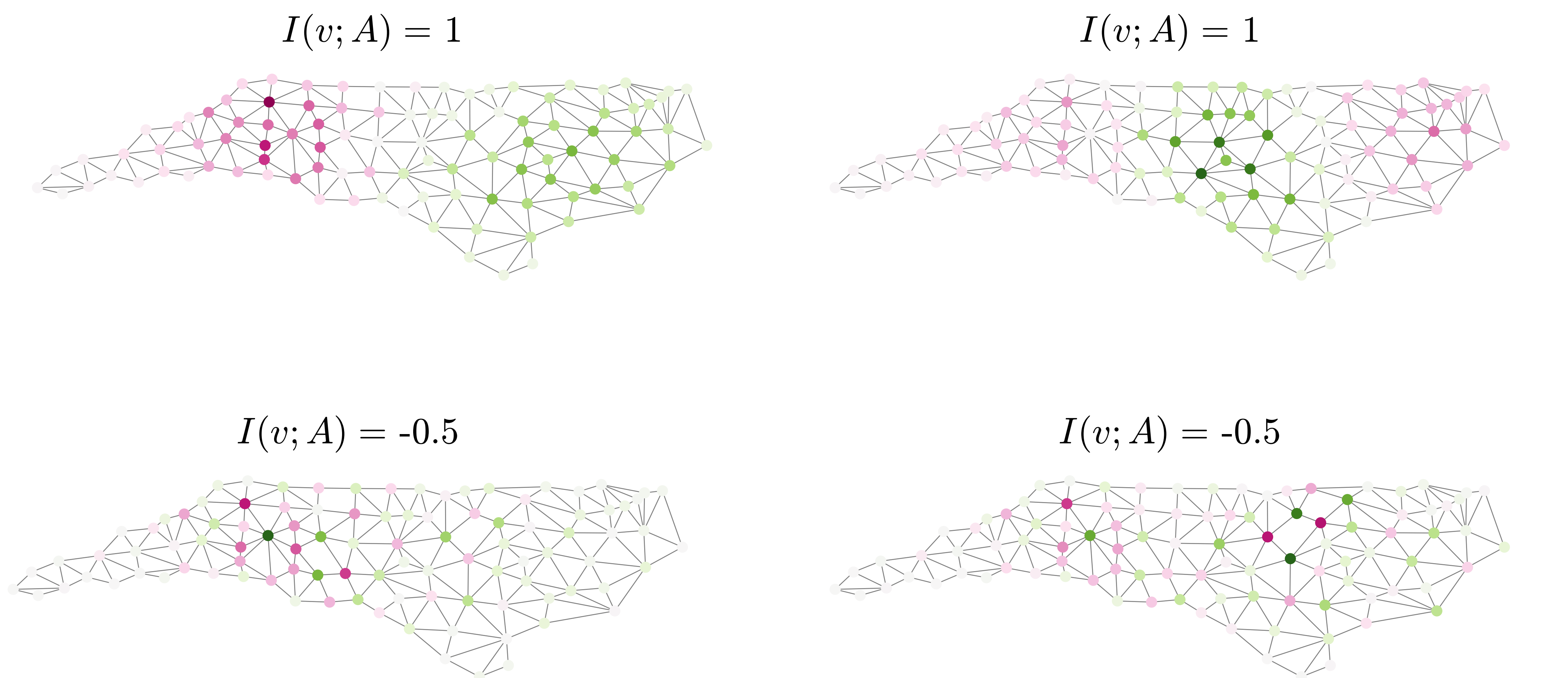
*(b)* $\mathtt{I}(\mathsf{v}; W) = \sum_{i=1}^{n-1} \alpha_i^2 \lambda_i \Big/ \sum_{i=1}^{n-1} \alpha_i^2$.

- This gives a decomposition tool to see what kinds of things make for large and small `I`.
- Also amenable to fast numerical computation in some cases.

Tufts UNIVERSITY

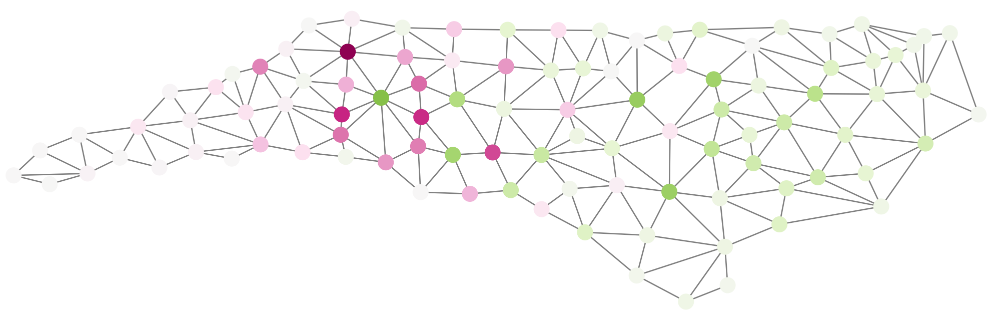
# Can We Compare with `I` ?

- Our spectral analysis allows us to answer the question: "are two functions on a network similar if they have similar `I` values"?
- Qualitatively yes, *if* `I` is very large or very small.

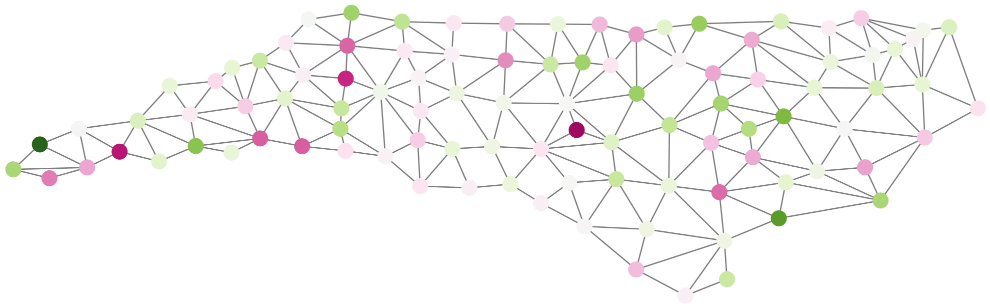


# Can We Compare with I ?

$I(v; A) = 0$



$I(v; A) = 0$



- When I is close to 0, "mixing" of different structures (e.g., clusters, localized checkerboards) makes inference difficult.

- Note: the spectral decomposition is what makes this "mixing" argument precise and practical.

Tufts UNIVERSITY

# Connection With Fourier Analysis

- By replacing the adjacency matrix with the *graph Laplacian*

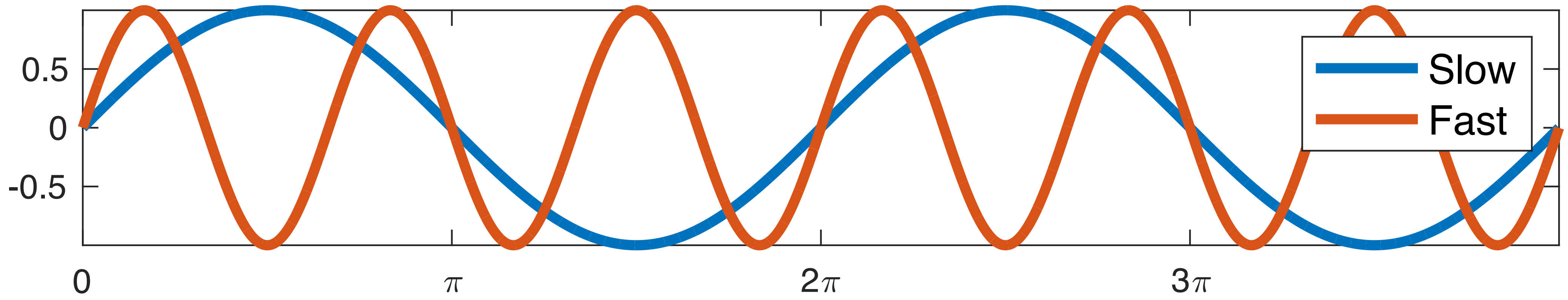
$$L = D - A \ (D_{ii} = \sum_{j=1}^{n} A_{ij},\ D_{ij} = 0 \text{ for } i \neq j)$$

we may interpret measures of spatial segregation in the context of *Fourier analysis* on graphs.

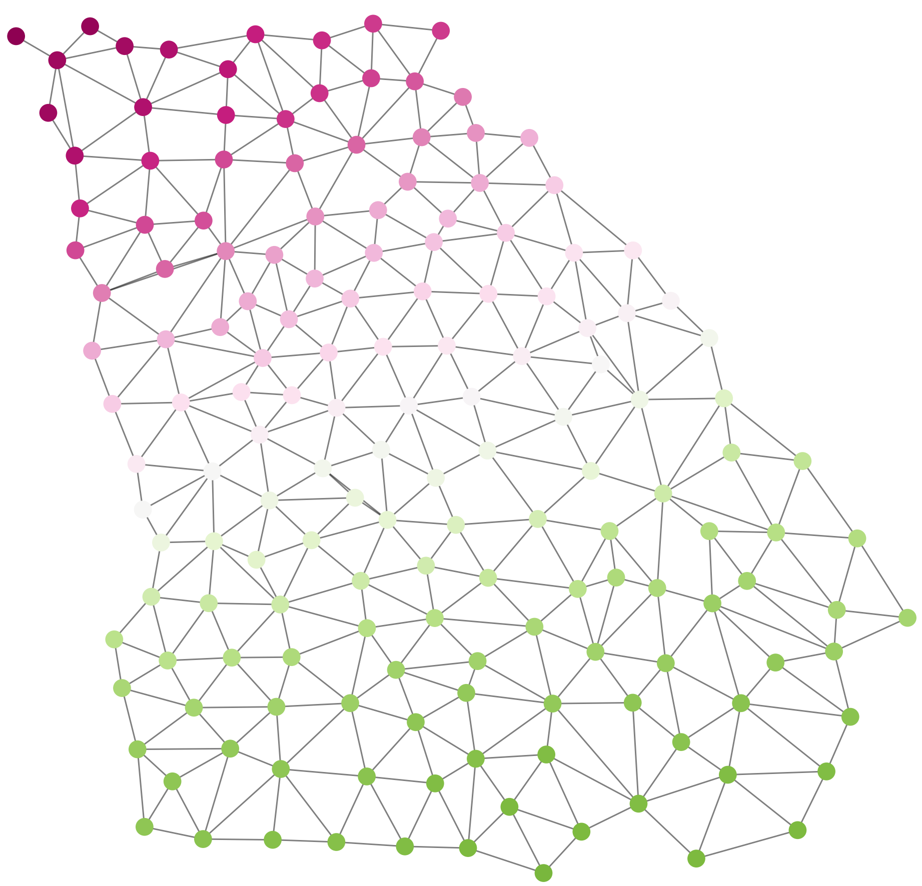
- Fourier analysis decomposes a function/signal into sines and cosines, capturing the *oscillatory structure* in the data.

Tufts UNIVERSITY

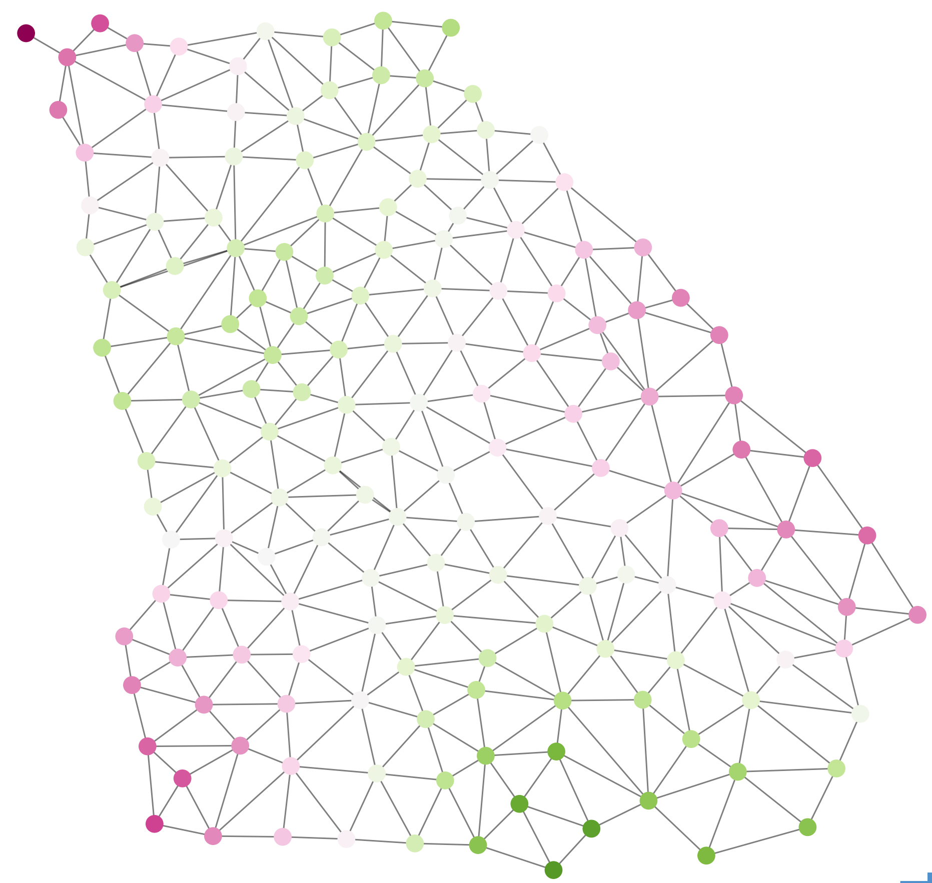
# Laplacian Oscillation on Graphs




Slow



Fast



Tufts
UNIVERSITY

# High Segregation <—> Low Energy

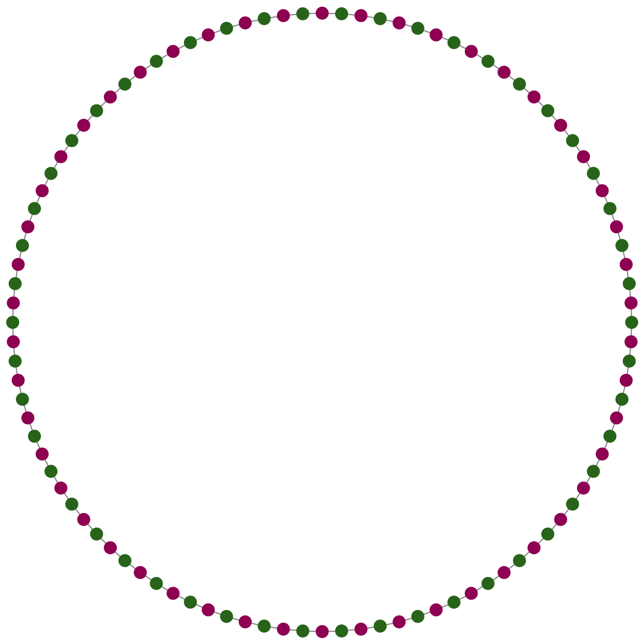
- Under this interpretation, highly segregated functions correspond to slowly oscillating Fourier modes, i.e. those with low energy.

- Highly anti-segregated functions (those with typical neighbor values different from themselves) are high energy.

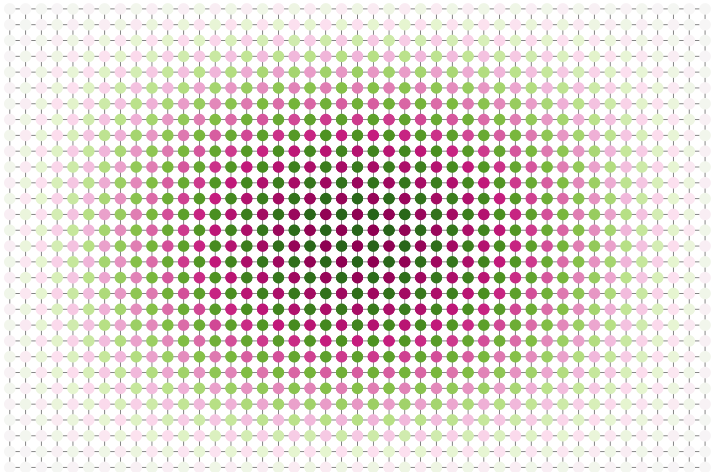
- The high energy interpretation does break down due to the irregularity of real geography graphs.

Tufts
UNIVERSITY

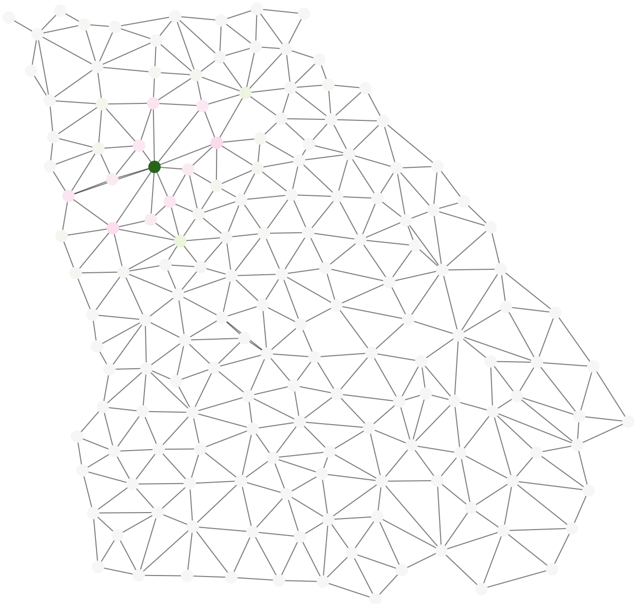
# High Energy and Localization



Highest Energy, Regular Graph



Highest Energy, Almost Regular Graph



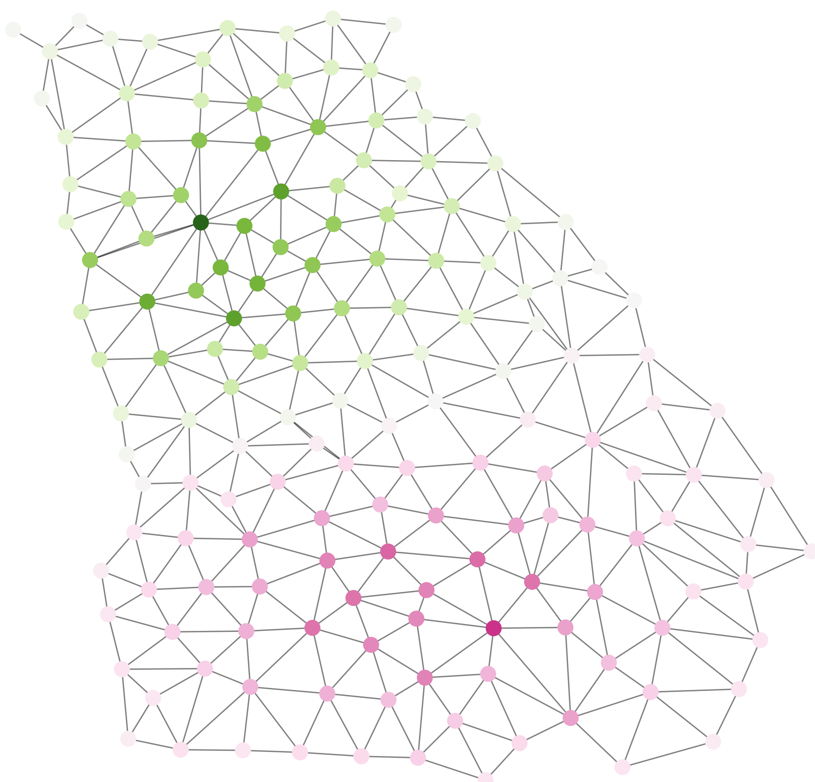
Highest Energy, Irregular Graph

- For highly regular graphs, there is a sense of "oscillation" in the highest frequency Fourier mode.
- Things are weirder in irregular graphs.
- Problem: characterize high-frequency functions on irregular graphs.

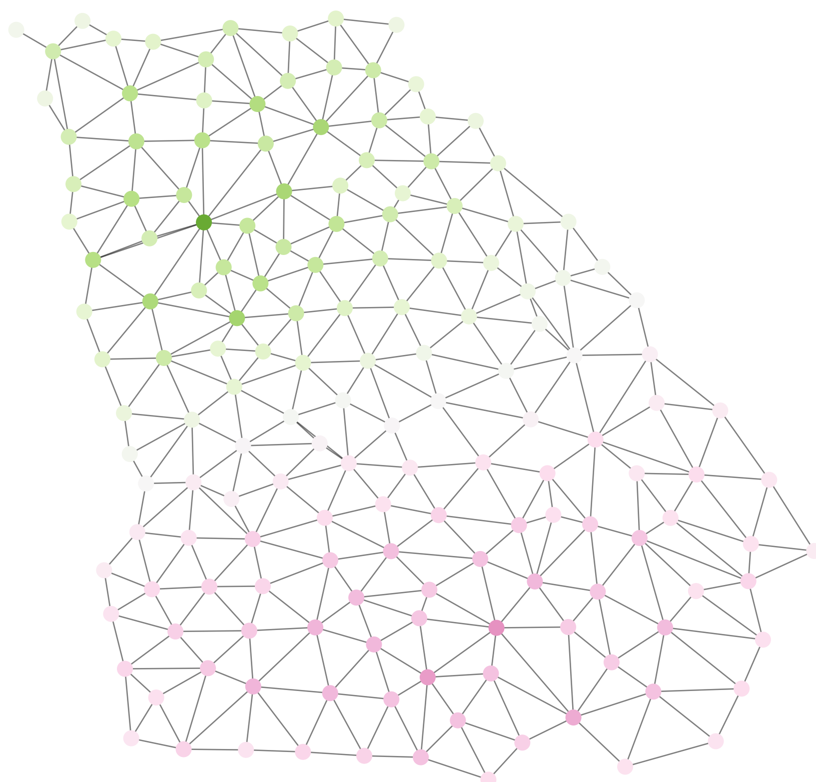
Tufts UNIVERSITY

# Random Walks and I

- One can imagine a random walk on the graph, where a walker must move to one of its neighbors (all with equal probability) at each time step.

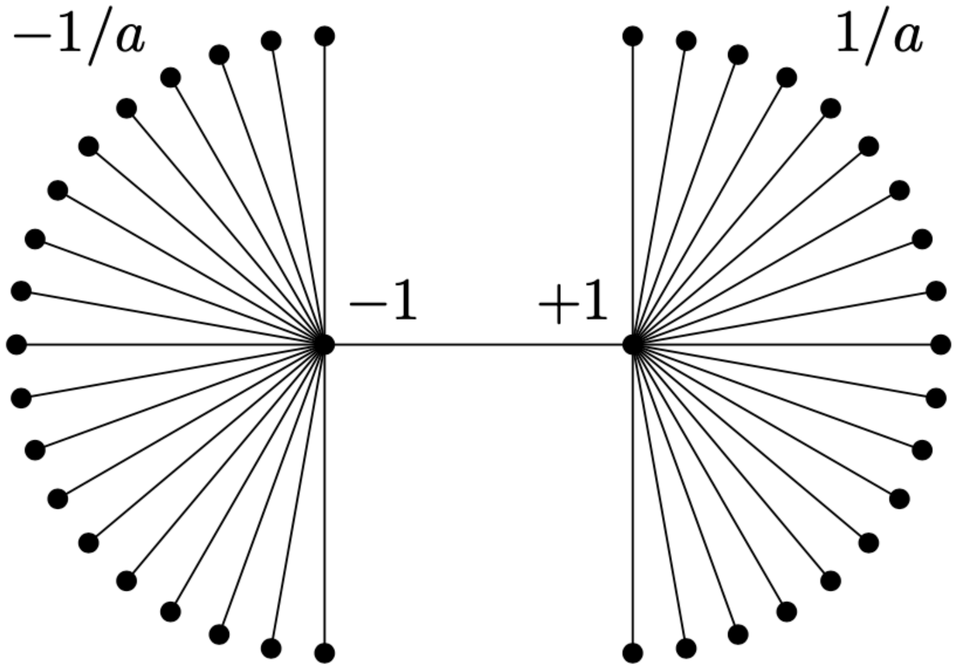


Start



After 10 Steps of Random Walk

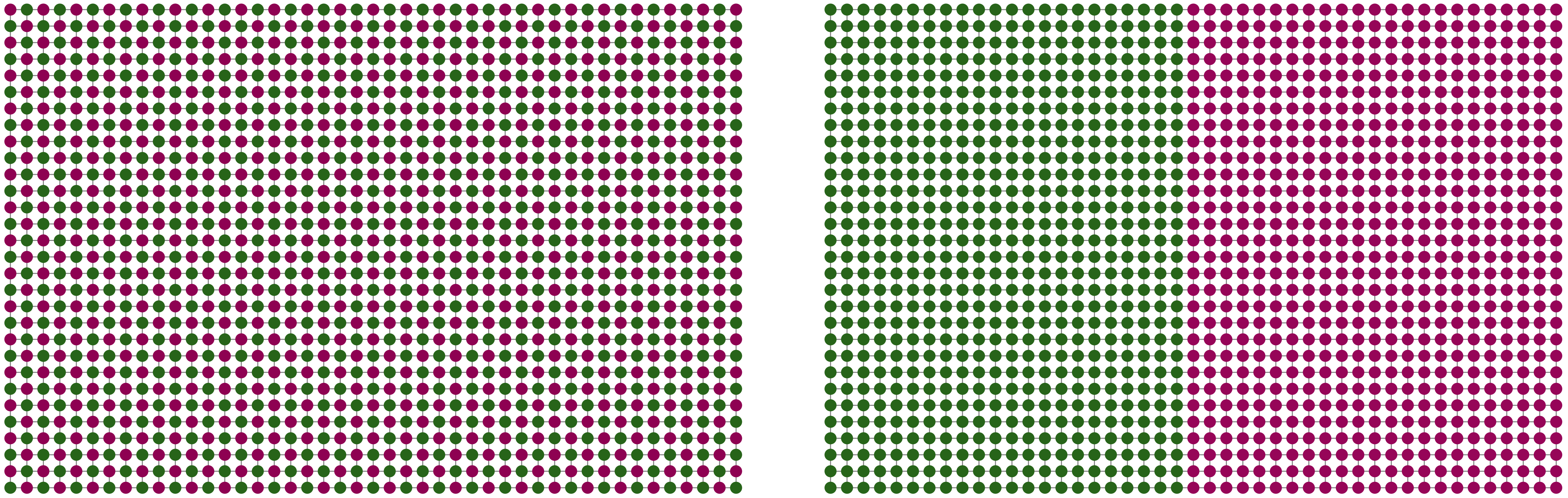
# Does Row-Normalizing Help?




- Let $P = D^{-1}A$ be the row-normalized random walk matrix.
- $\mathtt{I}(\cdot, P)$ still blows up on this class of graphs!
- Conventional wisdom around row-normalization is false.

# Random Walks and I

- This allows us to interpret I as a *correlation* across time steps.
- If there is a large degree of correlation across time, this indicates either very large or very small I values.



# (Bistochastic) Random Walks and $\mathtt{I}$

**Theorem.** *For a bistochastic matrix $Q$ and a column vector $\mathsf{v}$, consider $\mathsf{w} = \mathsf{v}^\top Q$, the value of $\mathsf{v}$ after one step of the Markov chain given by $Q$. Let $\sigma_0$ and $\sigma_1$ be the standard deviation of the values in $\mathsf{v}$ and $\mathsf{w}$ respectively, so that the ratio $\sigma_1/\sigma_0$ gives the variance reduction in one step of the walk. Let $\rho(\mathsf{v},\mathsf{w})$ be the correlation between the values in $\mathsf{v}$ and $\mathsf{w}$. Let $\mathsf{x} = \mathsf{v} - \bar{v}1$ and $\mathsf{y} = \mathsf{w} - \bar{w}1$ be the zero-centered vectors before and after applying $Q$. Then*

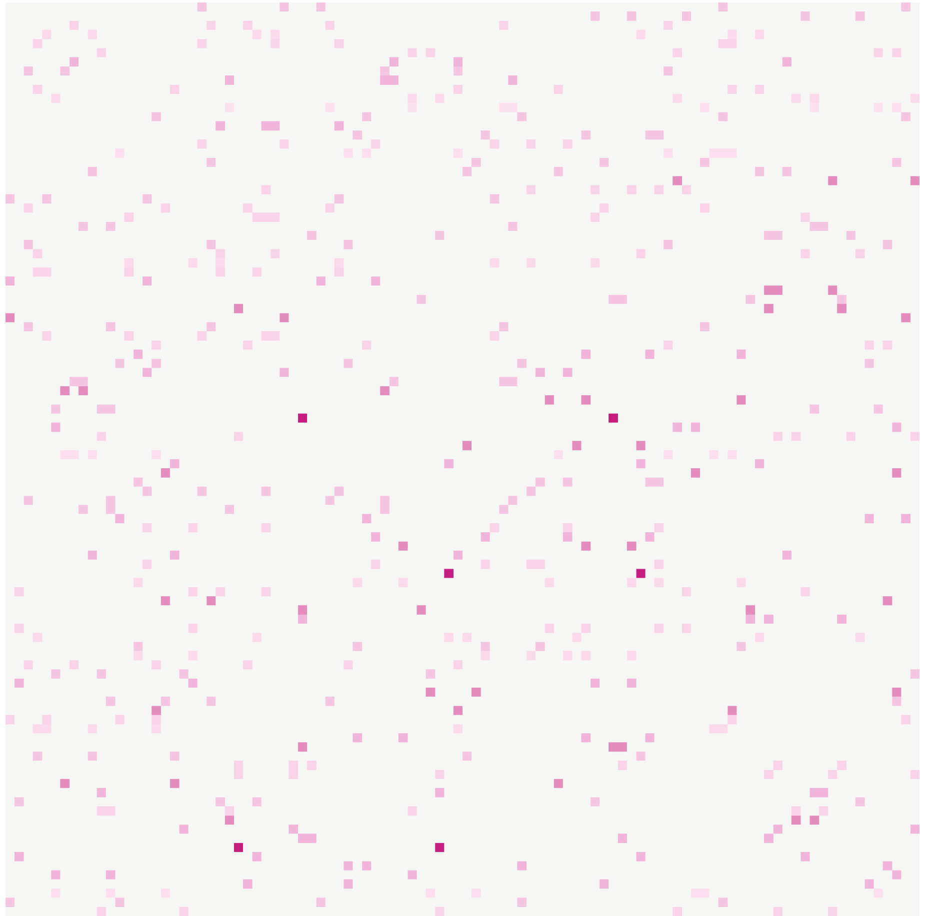
- $\mathtt{I}(\mathsf{v}; QQ^\top) = \left(\frac{\sigma_1}{\sigma_0}\right)^2$.
- $\mathtt{I}(\mathsf{v}; Q) = \frac{\mathsf{y}^\top \mathsf{x}}{\mathsf{x}^\top \mathsf{x}} = \rho(\mathsf{v},\mathsf{w}) \cdot \frac{\sigma_1}{\sigma_0}$.

- So, bistochastic matrices have a nice interpretation.
- They also address the unboundedness issues.
- Lazy choice we like: uniformizing Metropolis-Hastings matrix.

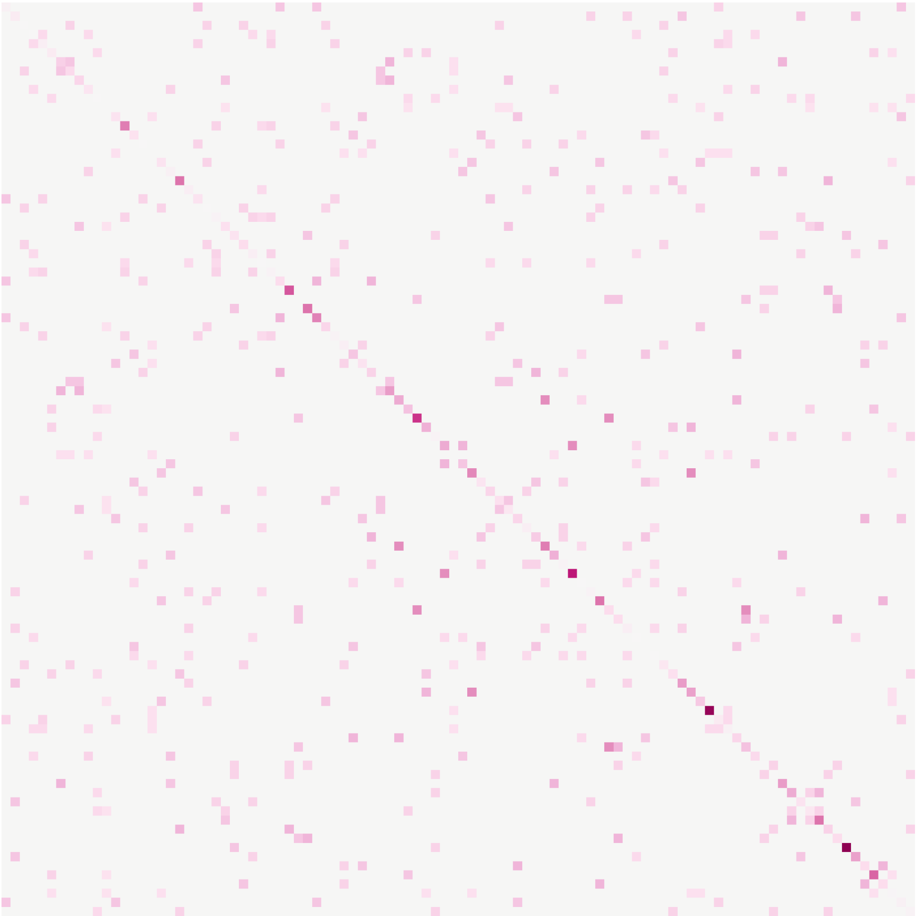
$$M_{ij} = \min\{P_{ij}, P_{ji}\},\ i \neq j$$

Tufts UNIVERSITY

# NC County Duals: $P$ v. $M$



$P$



$M$

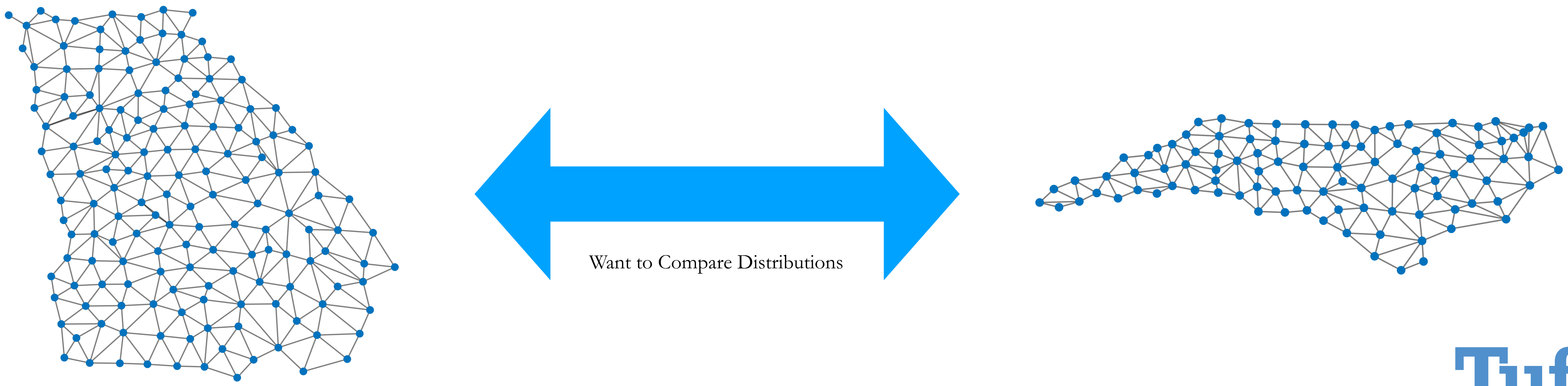
# Take Aways

- Graph theory (and linear algebra more generally) lets us investigate the properties of classical segregation measures.
- Commonly claimed properties of $\mathtt{I}$ do not hold in general, but only in highly structured cases.
- Comparing $\mathtt{I}$ within the same graph allows for certain qualitative inferences, but only in extreme cases.
- Interpretations in the language of Fourier analysis and random walks can illuminate.
- Maybe use $M$ ?

Tufts UNIVERSITY

# Optimal Transport on Graphs to Compare?

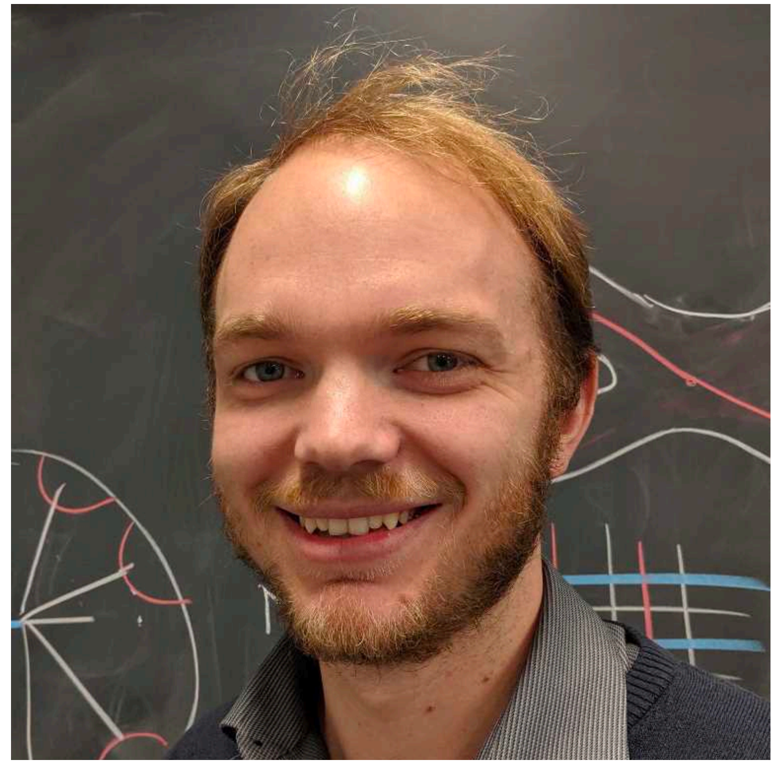
- Need to be able to compare *across graphs*.
- This can be formulated as a transport problem: map a distribution from one network to another in a cost minimizing way.
- Computational challenges, but has potential to allow for meaningful comparisons of communities on different graphs.




Tufts UNIVERSITY

# Collaborators, References, Acknowledgements

Moon Duchin, Tufts



Thomas Weighill, UNC Greensboro


- Duchin and Murphy. *Measuring Clustering and Segregation,* Political Geometry. Birkhäuser. 2022.
- Duchin, Murphy, and Weighill. *Measuring Segregation via Analysis on Graphs*. SIAM Journal on Matrix Analysis and Applications (to Appear). arXiv:2112.10708



NSF

DMS 1912737
DMS 1924513
CCF-1934553

The Camille & Henry Dreyfus Foundation

DISC

Tufts University