# Graph and Networks: Discrete Approaches to Redistricting

Daryl DeFord

MIT/Tufts
Metric Geometry and Gerrymandering Group

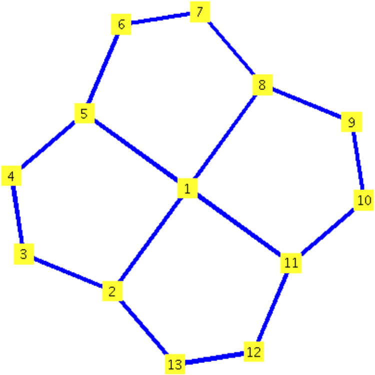
Voting Rights Data Institute
June 18, 2019

# Outline

# Graph Theory

## Definition (Graph)

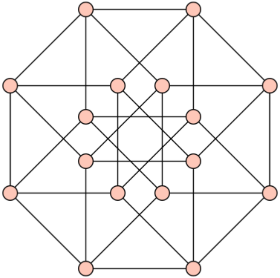
A Graph $G = (V, E)$ is a set of nodes $V$ and a set of edges $E \subseteq V \times V$.

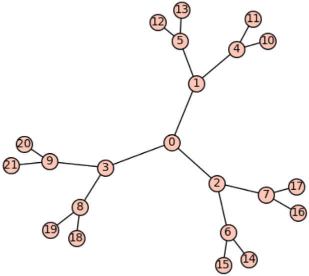


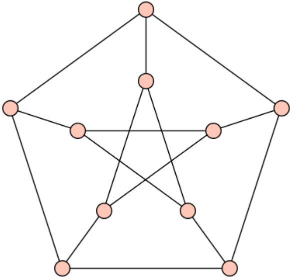
# Graph Questions

- Are there Hamiltonian/Eulerian Paths? If so, how many?
- Are there perfect matchings? If so, how many?
- Is it possible color the nodes with $k$ colors so that no neighboring nodes are the same color?
- What is the largest set of nodes that share no edges?
- How many edges/vertices must be removed to disconnect the graph?
- Is it possible to embed the graph in the plane without any edges crossing?
- How many spanning trees does the graph have?
- What is the automorphism group of the graph?
- ...

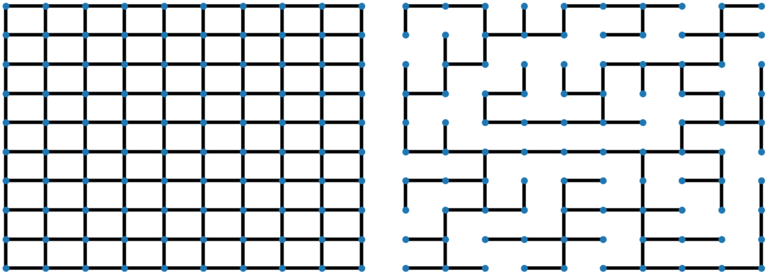
# Graph Examples





# Spanning Trees



# Graph Theorems

**Theorem**

*Every graph has an even number of odd-degree vertices*

**Theorem**

*The number of perfect matchings in an $m \times n$ grid graph is:*

$$\sqrt[4]{\prod_{j=1}^{m}\prod_{k=1}^{n}\left(4\cos^2\left(\frac{\pi j}{m+1}\right)+4\cos^2\left(\frac{\pi k}{n+1}\right)\right)}$$

**Theorem**

*Every graph where every node has even degree has an Eulerian path.*

**Theorem**

*Determining whether an arbitrary graph has a Hamiltonian cycle is NP–Hard*

# What is a social network?

**Definition (Social Network)**

Mathematically, a social network is represented by a collection of “nodes” representing individual actors and a set of “edges” representing a binary relationship between the actors.

# What is a social network?

## Definition (Social Network)

Mathematically, a social network is represented by a collection of "nodes" representing individual actors and a set of "edges" representing a binary relationship between the actors.

## Example

What kinds of systems can social networks describe?

- What could be represented by nodes?

# What is a social network?

## Definition (Social Network)

Mathematically, a social network is represented by a collection of "nodes" representing individual actors and a set of "edges" representing a binary relationship between the actors.

## Example

What kinds of systems can social networks describe?

- What could be represented by nodes?
    - **Academic Departments**

# What is a social network?

## Definition (Social Network)

Mathematically, a social network is represented by a collection of “nodes” representing individual actors and a set of “edges” representing a binary relationship between the actors.

## Example

What kinds of systems can social networks describe?

- What could be represented by nodes?
  - **Academic Departments**
- What type of edges could connect them?

# What is a social network?

## Definition (Social Network)

Mathematically, a social network is represented by a collection of "nodes" representing individual actors and a set of "edges" representing a binary relationship between the actors.

## Example

What kinds of systems can social networks describe?

- What could be represented by nodes?
  - **Academic Departments**
- What type of edges could connect them?
  - **Located in same building**
  - **Students who major in both**
  - **Crosslisted courses**

# What is a social network?

## Definition (Social Network)

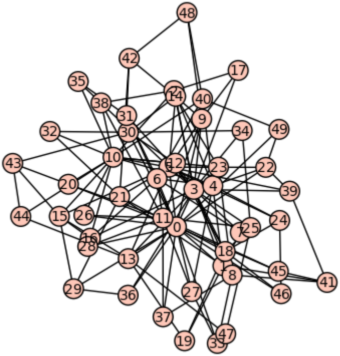
Mathematically, a social network is represented by a collection of "nodes" representing individual actors and a set of "edges" representing a binary relationship between the actors.

## Example

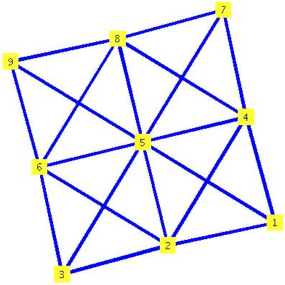
What kinds of systems can social networks describe?

- What could be represented by nodes?
  - **VRDI Students**
- What type of edges could connect them?
  - **Same Dorm Room**
  - **Facebook friends**
  - **Speak at least twice a week**

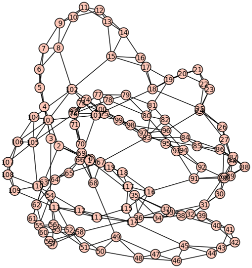
# Social Networks



# Examples

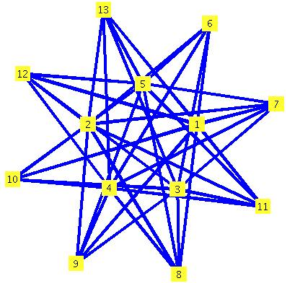



(a) Graph

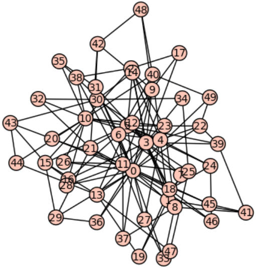


(b) Network

# Examples

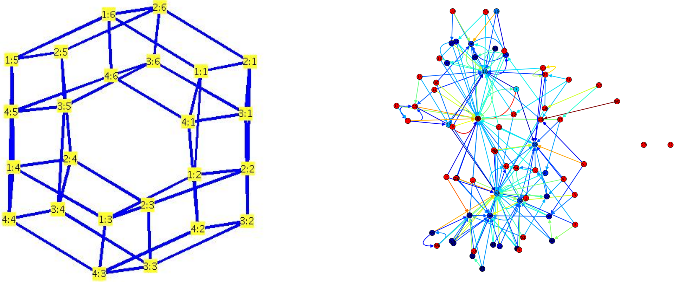



(a) Graph



(b) Network

# Examples




(a) Graph

(b) Network

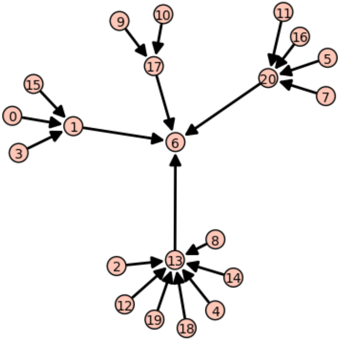
# Comparing Graphs to Networks

- Annotations
- Topology

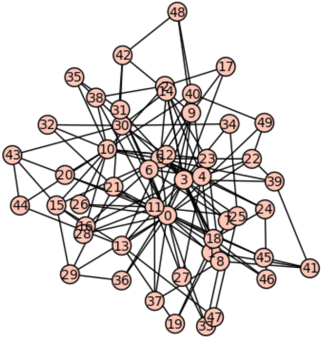
# Comparing Graphs to Networks

- Annotations
- Topology
- Fuzziness of data
- Errors in collection
- People are complicated
- Most importantly: **Different relevant questions**
  - Centrality
  - Community Structure
  - “Social” Dynamics

# Centrality



# Centrality


48
42
17
35
14
40
38
31
9
30
34
49
32
43
10
12
23
22
6
3
4
39
20
21
44
15
26
11
0
24
25
16
28
18
13
45
8
41
29
36
27
46
37
19
47

# Centrality

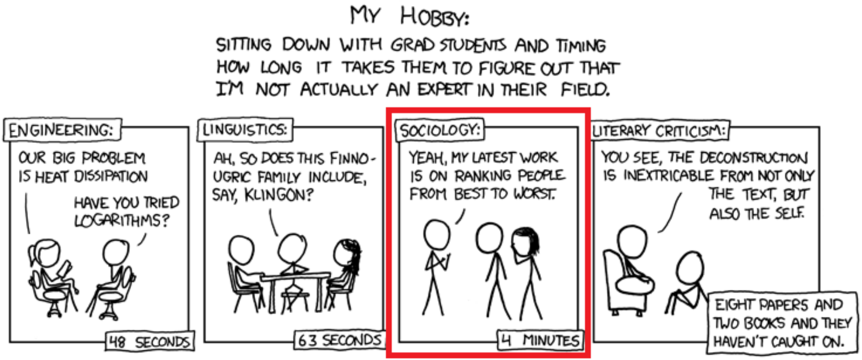
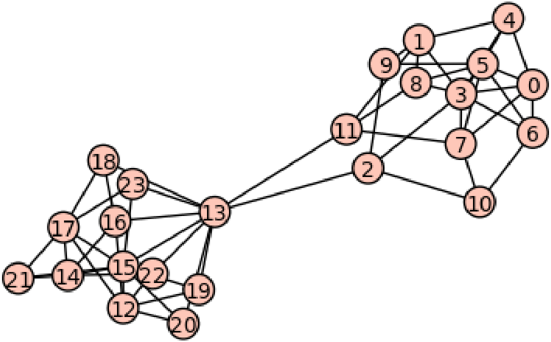


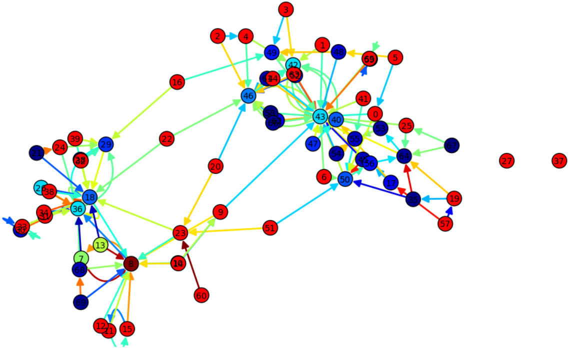

Figure: Relevant comic by Randall Munroe[1]

[1] https://xkcd.com/451/

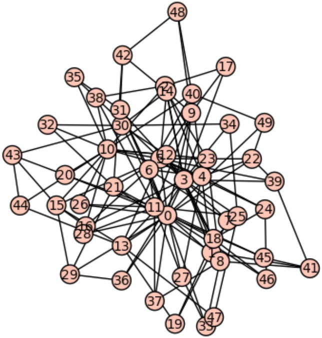
# Clustering



# Clustering



# Clustering


48
42
17
35
14
40
38
31
9
32
30
34
49
43
10
12
23
22
6
3
4
39
20
21
44
15
26
11
0
24
25
16
28
18
13
45
41
8
29
36
27
46
37
47
19
33

# Common Properties of Social Networks

**Example (What features distinguish social networks?)**

- ?

# Common Properties of Social Networks

**Example (What features distinguish social networks?)**

- ?
- Transitivity
- Community structure
- Small average path length
- Long–tailed degree distribution
- Hubs
- ...

# Ego Networks

**Definition (Ego Network)**

An ego network is a social network centered at a particular individual containing their connections and the connections between their "friends."

# Ego Networks

**Definition (Ego Network)**

An ego network is a social network centered at a particular individual containing their connections and the connections between their "friends."

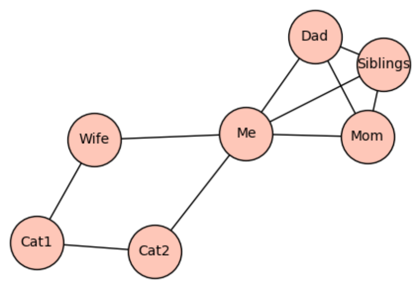
**Example (Draw your ego network)**

# Ego Networks

## Definition (Ego Network)

An ego network is a social network centered at a particular individual containing their connections and the connections between their "friends."

## Example (Draw your ego network)



# How to construct networks?

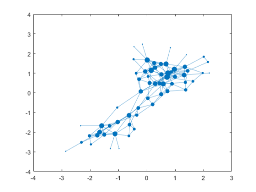
**Example (Which edges to add?)**

- ?

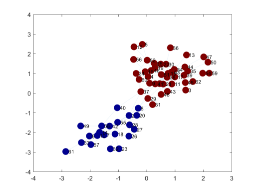
# How to construct networks?

## Example (Which edges to add?)

- ?
- Proximity




(a) Centrality

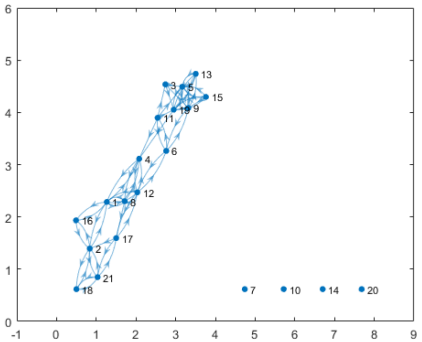



(b) Clustering

Figure: Dolphin social network[1]

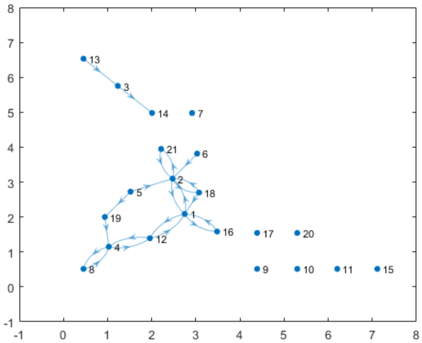
s [1] D. Lusseau, K. Schneider, O. Boisseau, Patti Haase, E. Slooten, and S. Dawson, The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations, Behavioral Ecology and Sociobiology 54 (2003), no. 4, 396–405.

# Different Perspectives on “Friendship”



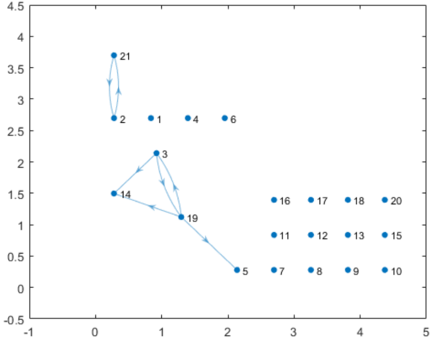

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



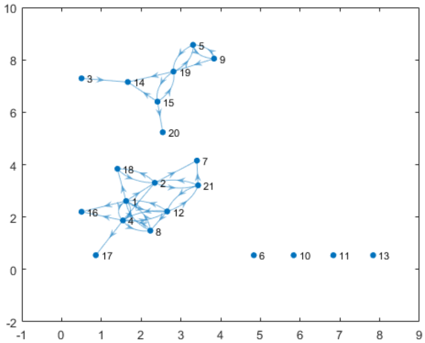

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on "Friendship"



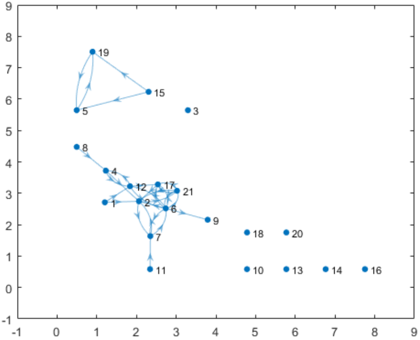

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



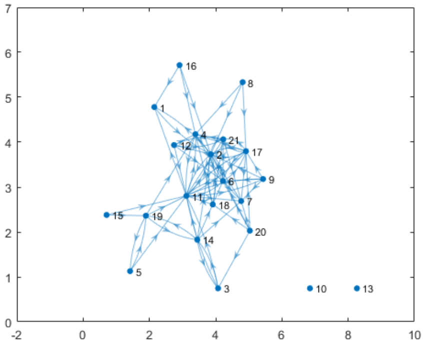

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



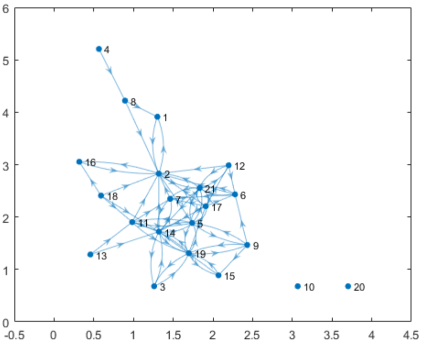

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



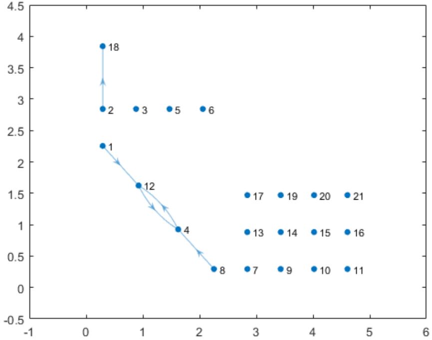

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



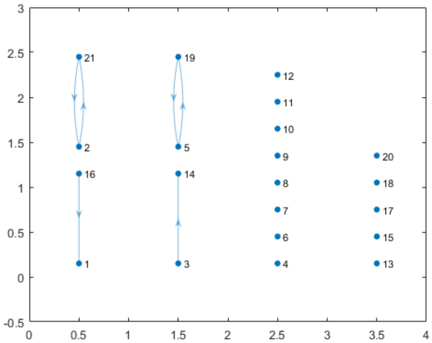

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



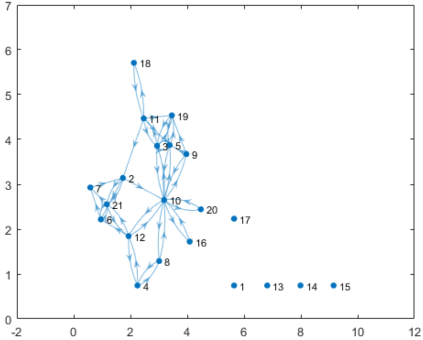

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



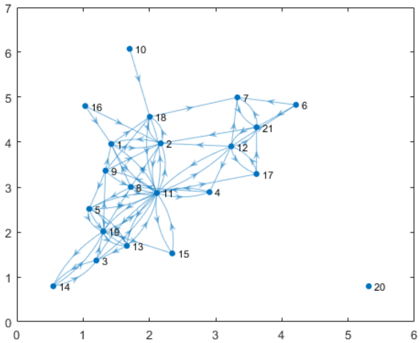

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



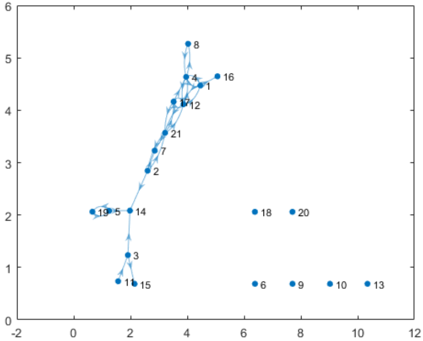

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



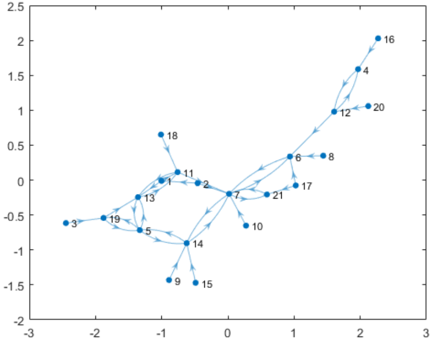

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



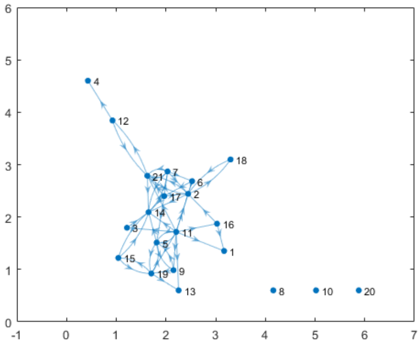

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



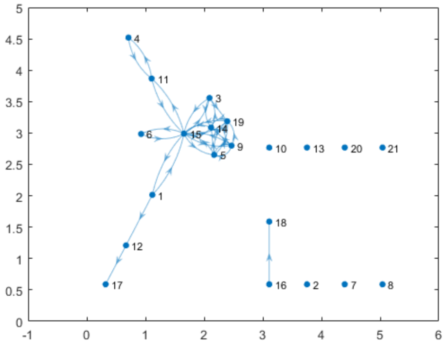

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



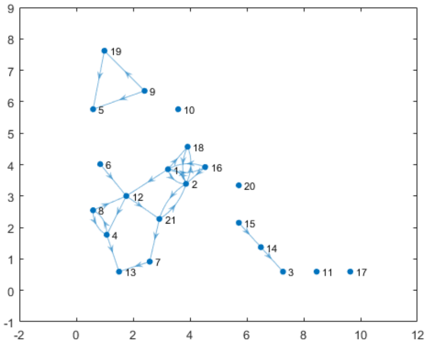

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



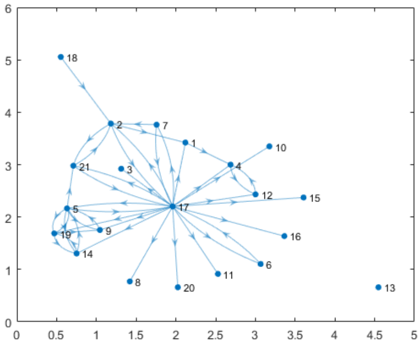

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



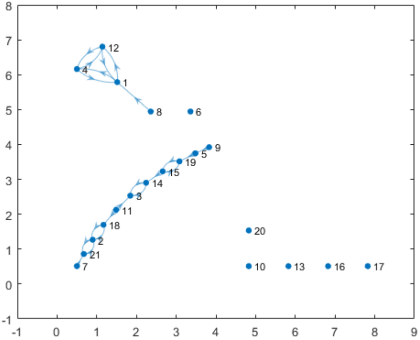

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



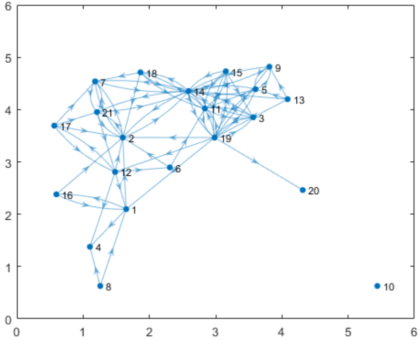

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



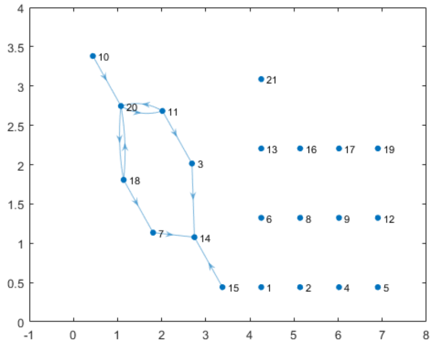

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on "Friendship"



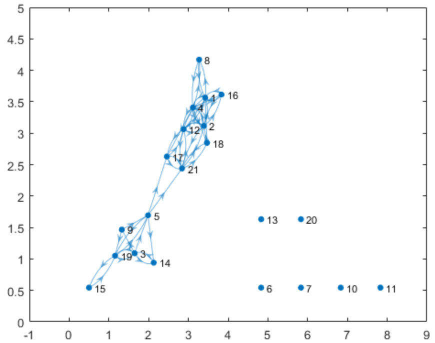

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



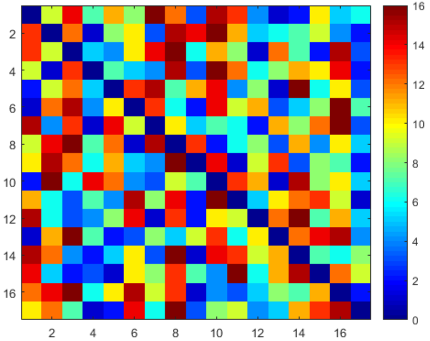

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# Different Perspectives on “Friendship”



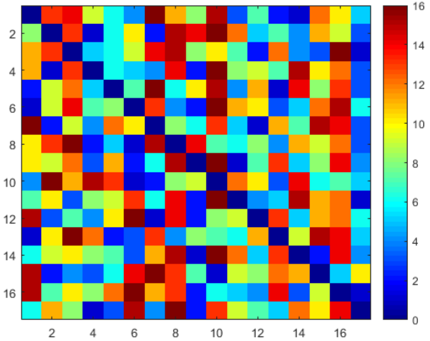

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

# "Friendship" over Time



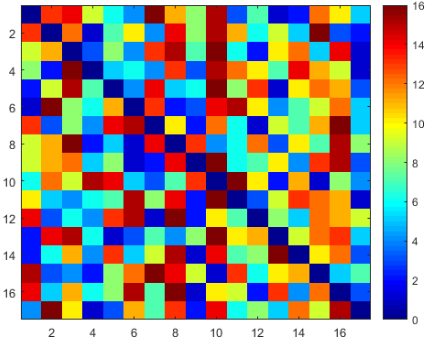

Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

# "Friendship" over Time



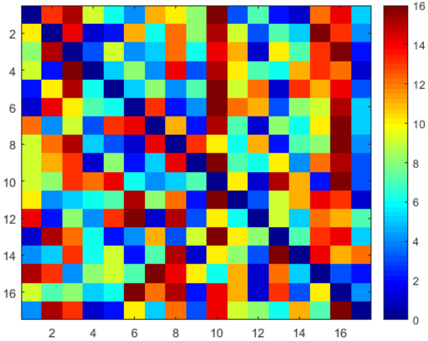

Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

# "Friendship" over Time



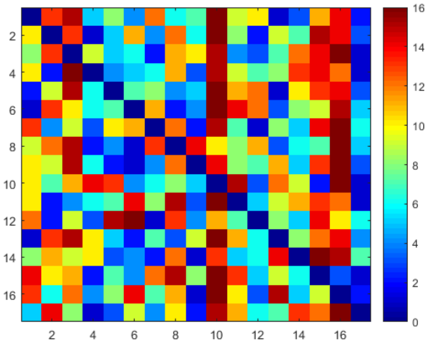

Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

# “Friendship” over Time



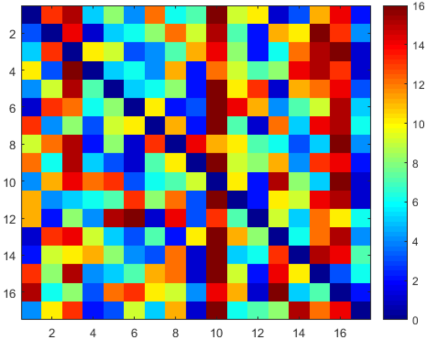

Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

# "Friendship" over Time



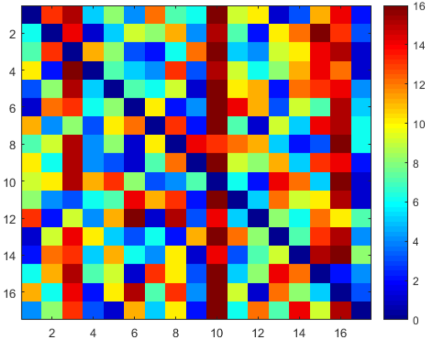

Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

# “Friendship” over Time



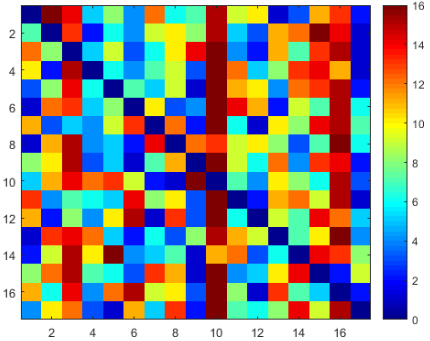

Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

# "Friendship" over Time



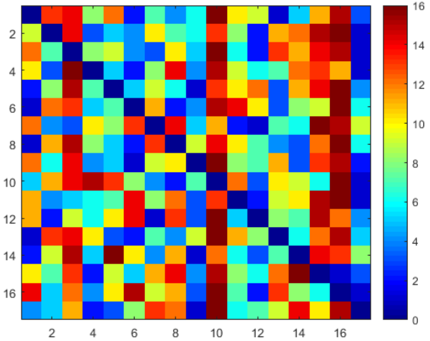

Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

# "Friendship" over Time



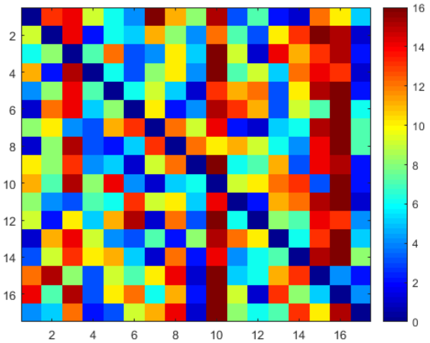

Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

# "Friendship" over Time



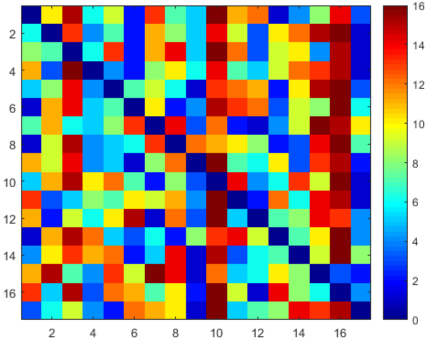

Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

# “Friendship” over Time



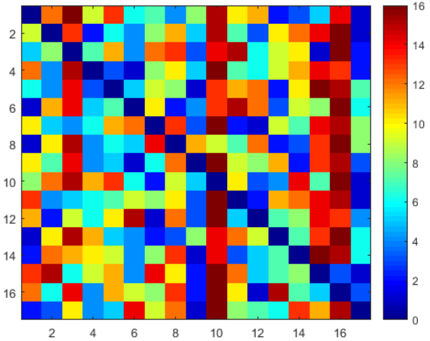

Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

# "Friendship" over Time



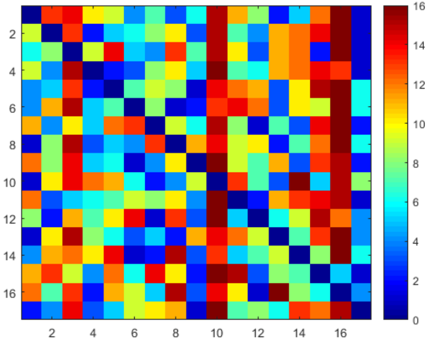

Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

# “Friendship” over Time



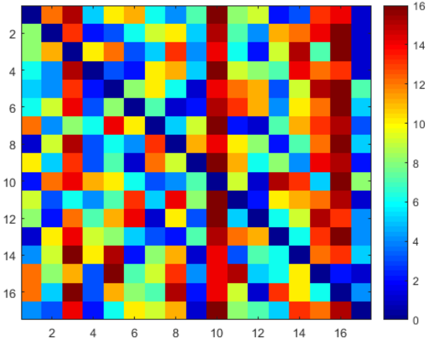

Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

# "Friendship" over Time




Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

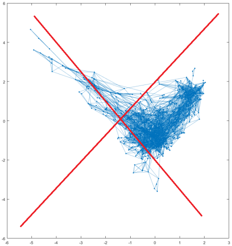
# "Friendship" over Time

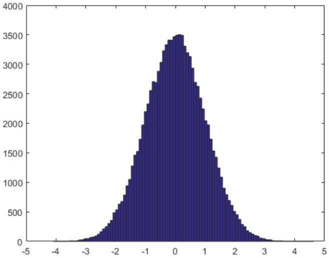



Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

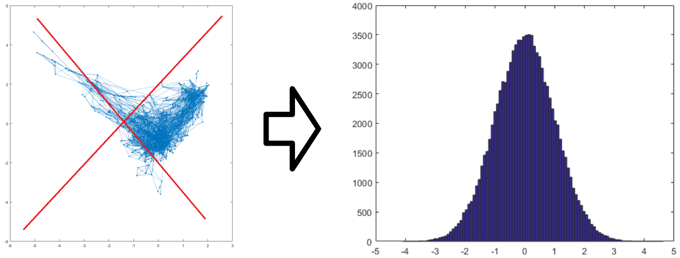
# Random Networks

# Random Networks





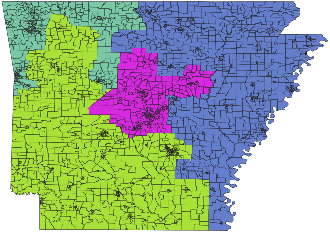
# Random Networks

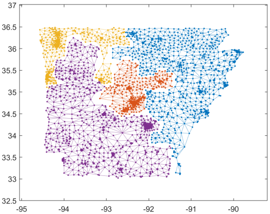



## Definition (Null Model)

A random network, parameterized to match some features of a given network, used to compare "expected" network measures.

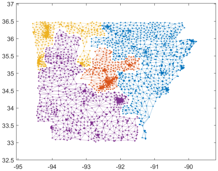
# Dual Graphs

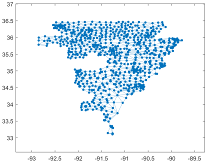


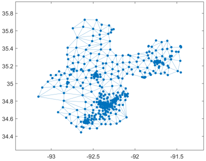


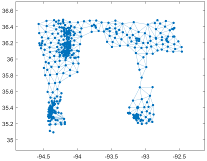

In order to study this problem mathematically we need to abstract the process of districting into the realm of mathematical objects. The first step is to discretize!

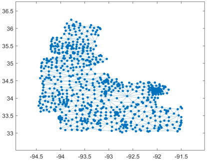
# Graph Partitioning



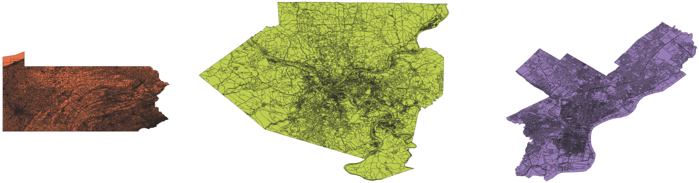



# Levels of Data Resolution

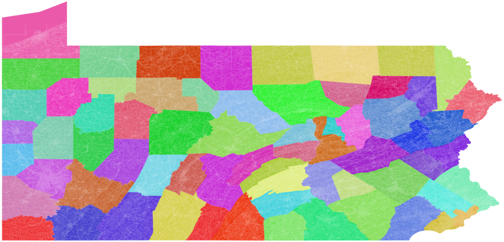
- Blocks
- Block Groups
- Tracts
- Precincts
- Wards
- Municipalities
- Counties

# Blocks

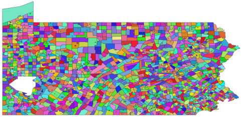


(a) Pennsylvania (b) Allegheny (c) Philadelphia

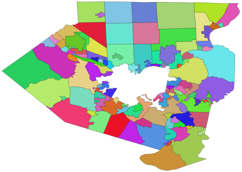
# Counties



# Municipalities

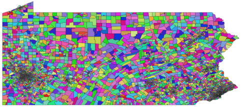


(a) Pennsylvania

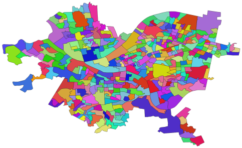


(b) Allegheny

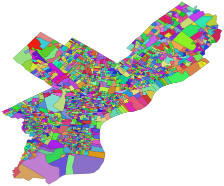
# Precincts



(a) Pennsylvania

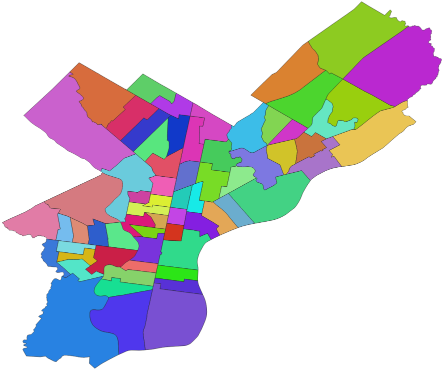


(b) Pittsburgh

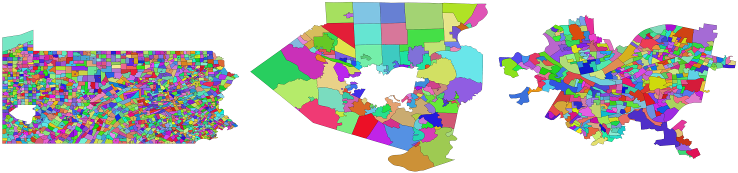


(c) Philadelphia

# Wards



# Putting Them Together

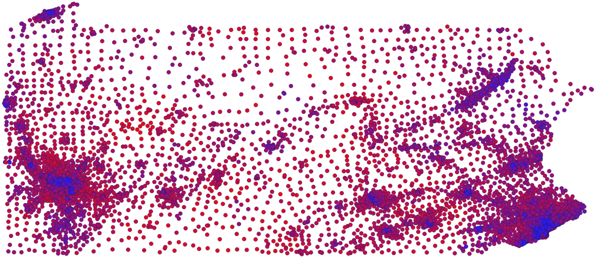


(a) Pennsylvania (b) Allegheny (c) Pittsburgh

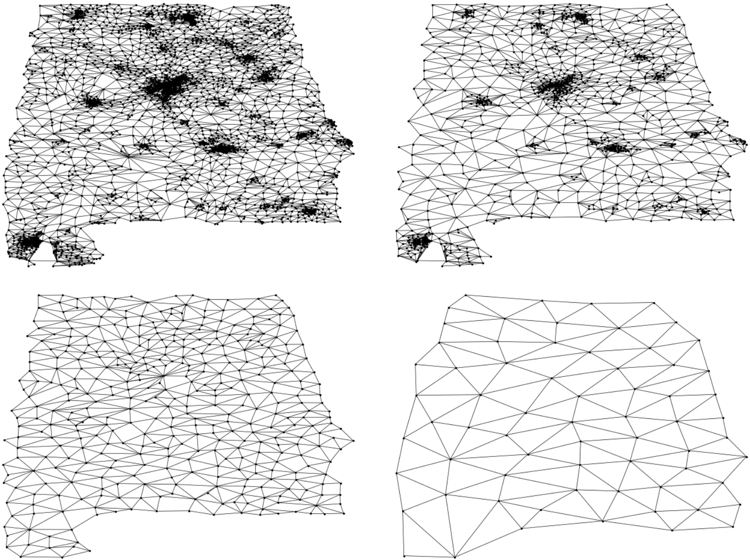
# Putting Them Together

(a) Blocks (b) Precincts (c) Wards

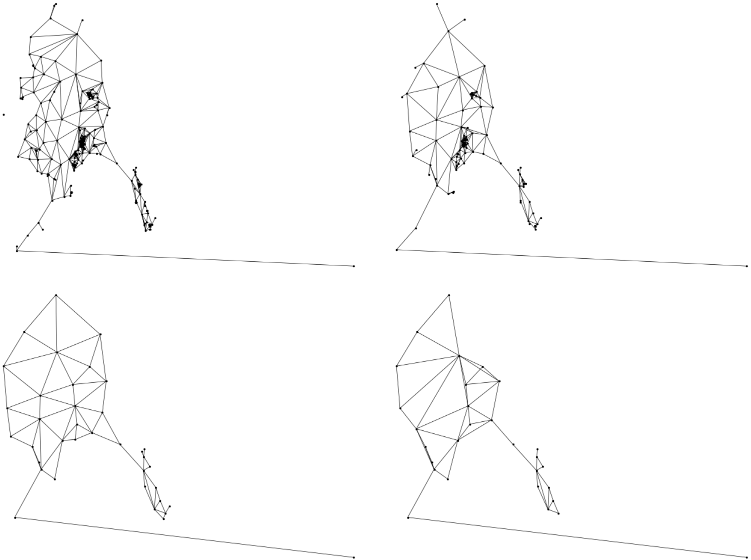
# Partisanship Measures



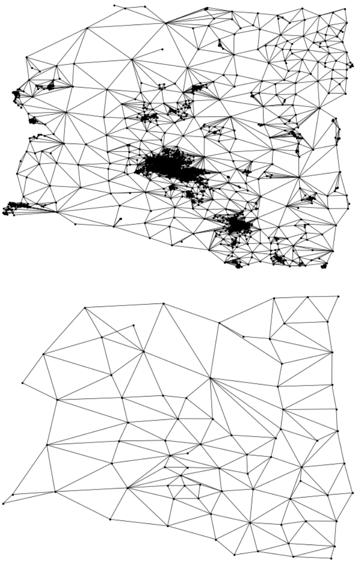
# Census Dual Graphs



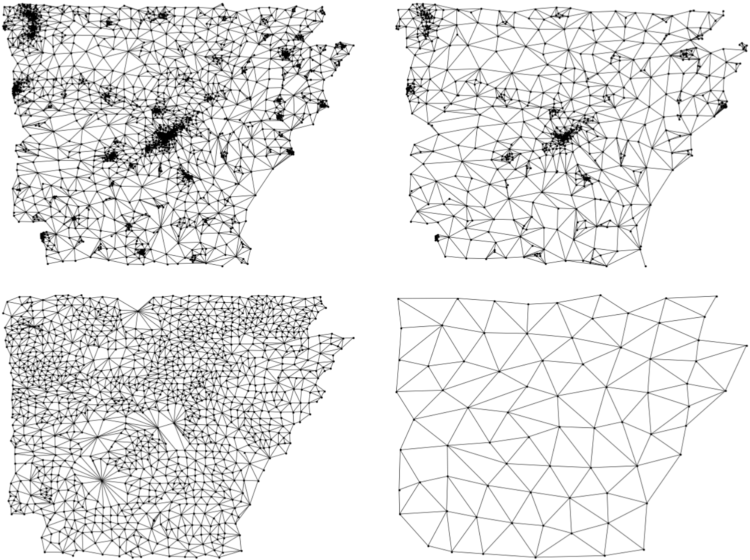
# Census Dual Graphs



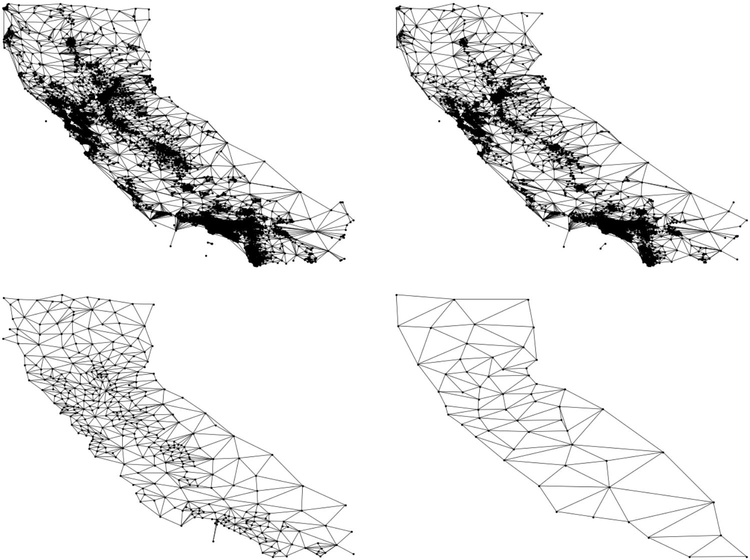
# Census Dual Graphs



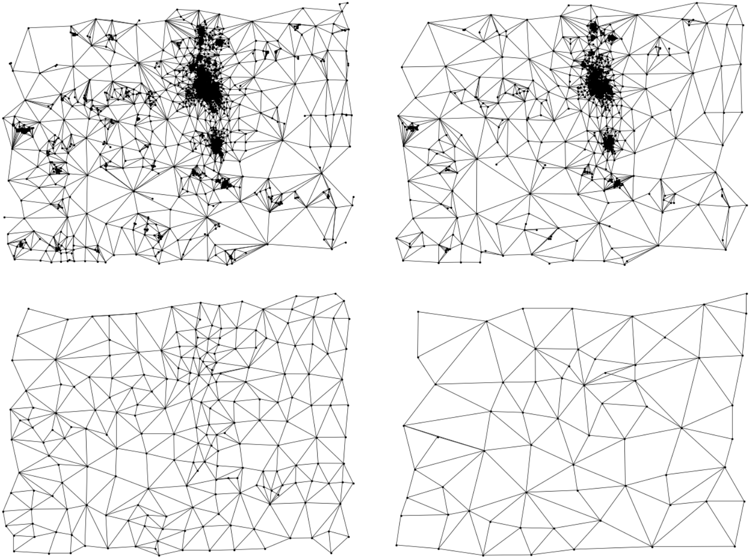
# Census Dual Graphs



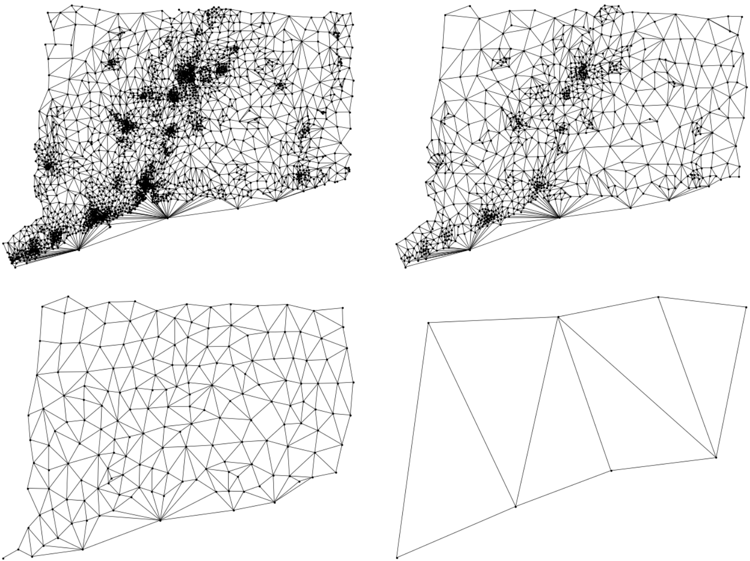
# Census Dual Graphs



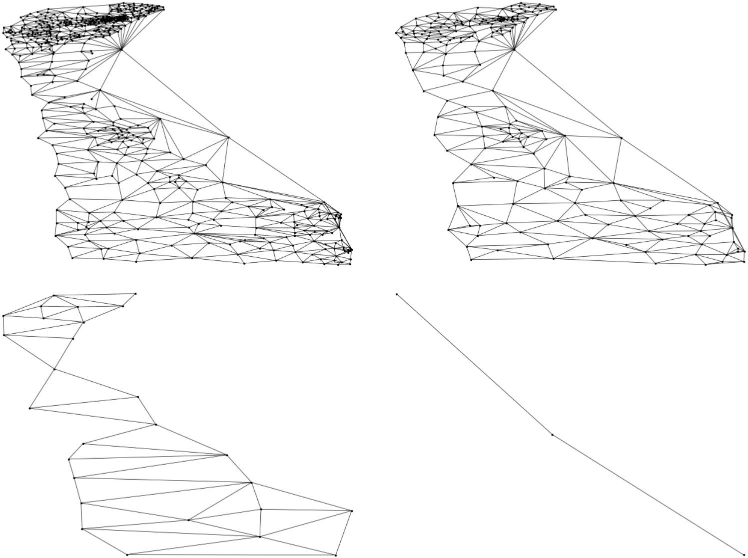
# Census Dual Graphs



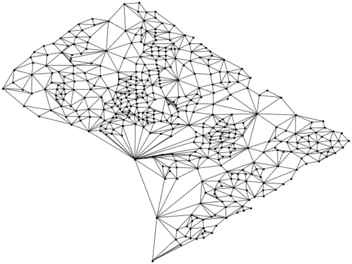
# Census Dual Graphs

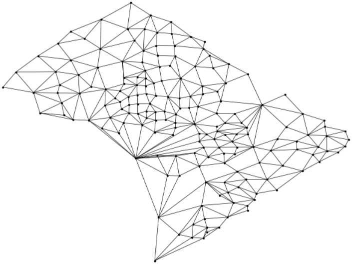


# Census Dual Graphs

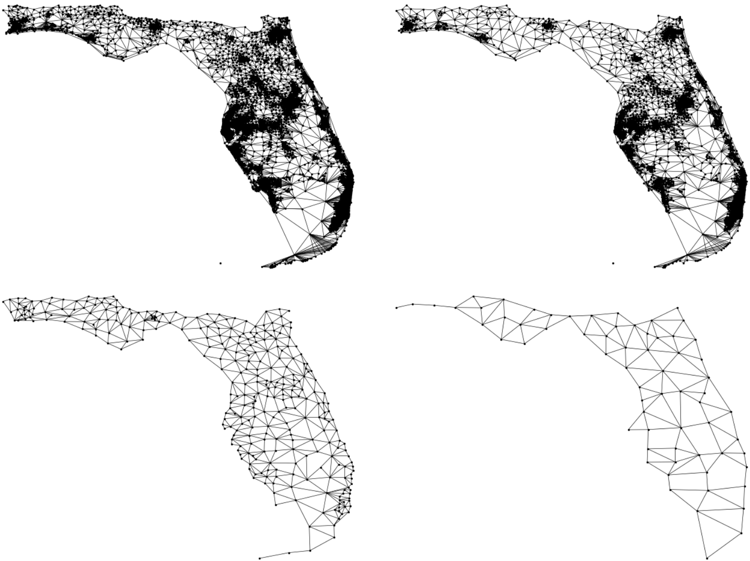


# Census Dual Graphs

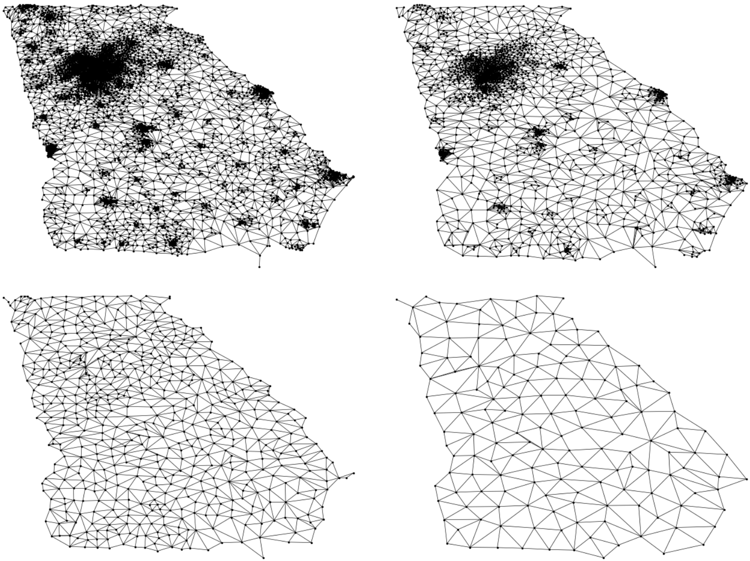




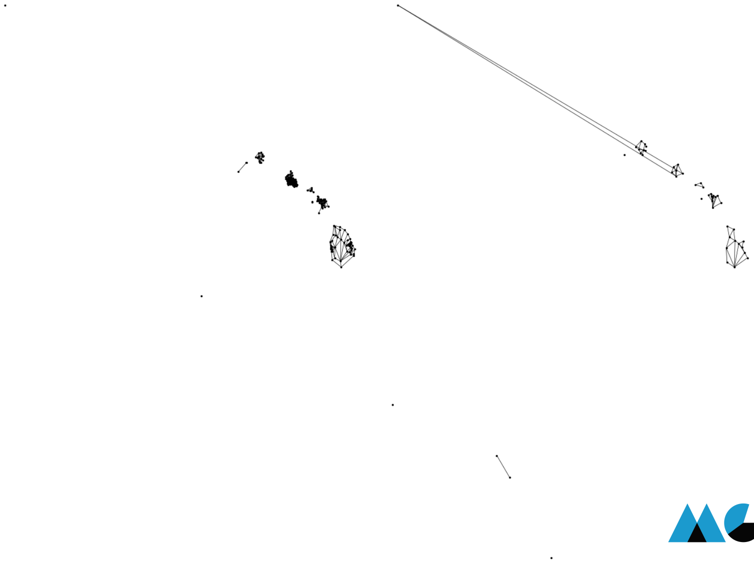
# Census Dual Graphs



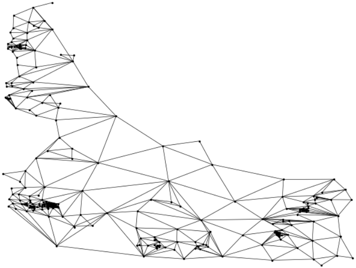
# Census Dual Graphs

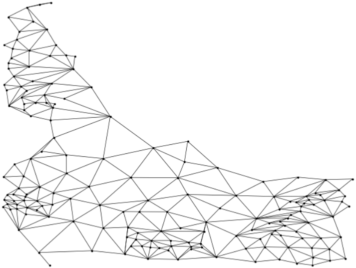


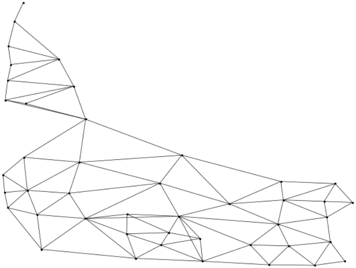
# Census Dual Graphs



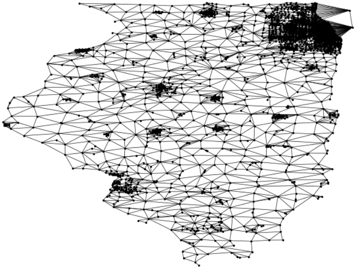
# Census Dual Graphs

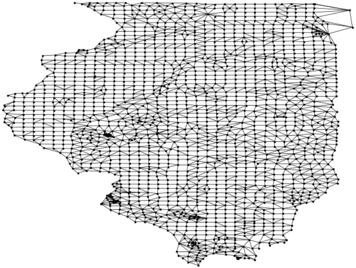


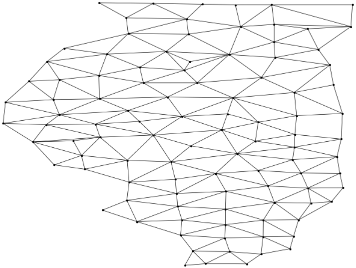




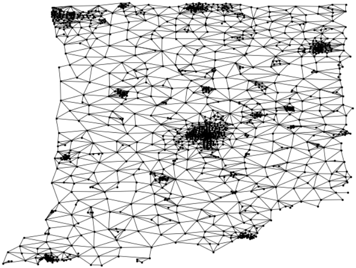
# Census Dual Graphs

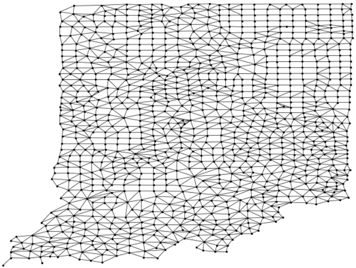


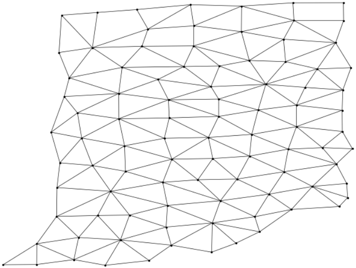




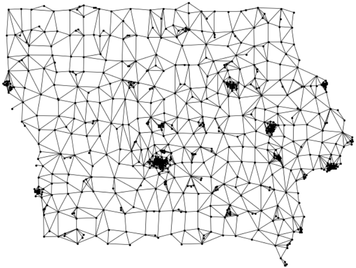
# Census Dual Graphs

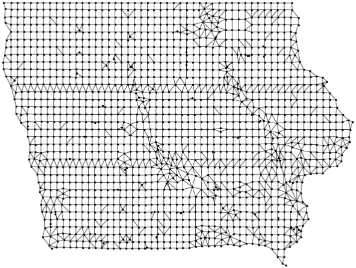


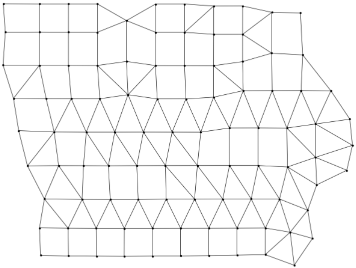




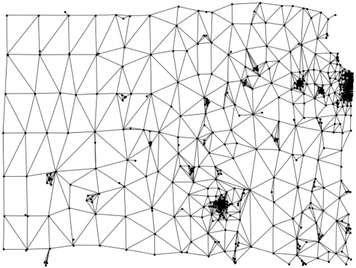
# Census Dual Graphs

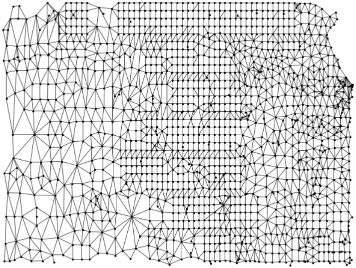


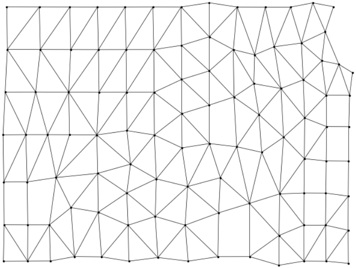




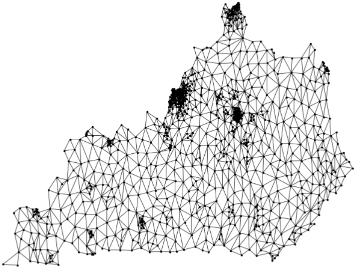
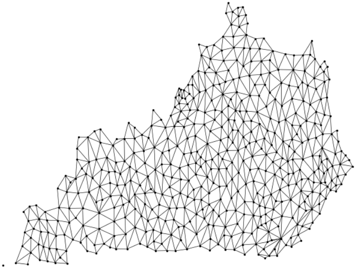
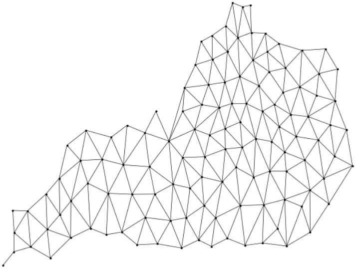
# Census Dual Graphs



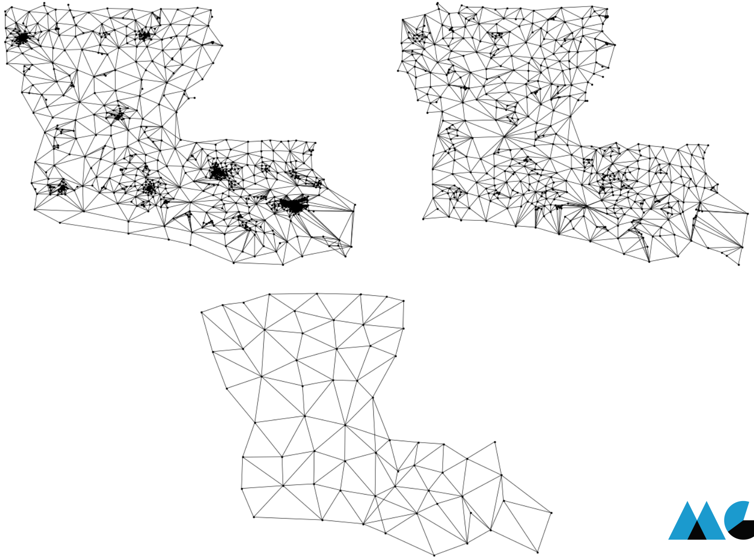




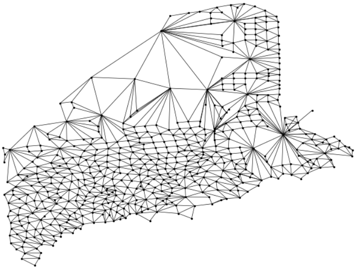
# Census Dual Graphs

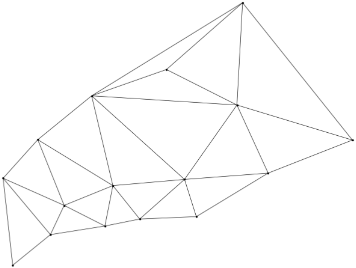




# Census Dual Graphs

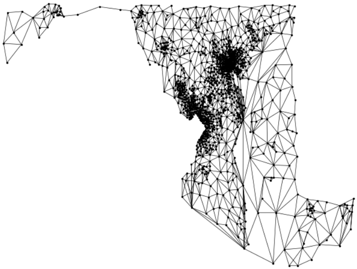


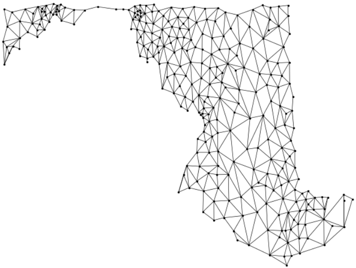
# Census Dual Graphs

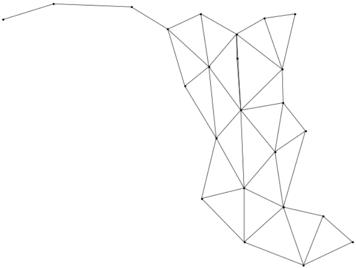




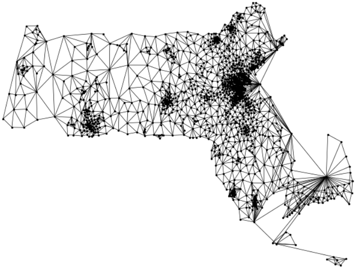
# Census Dual Graphs

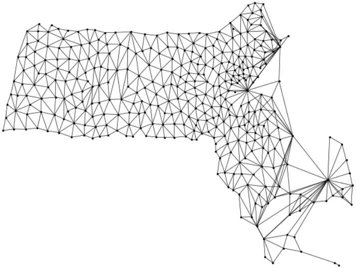


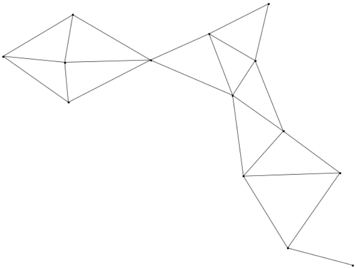




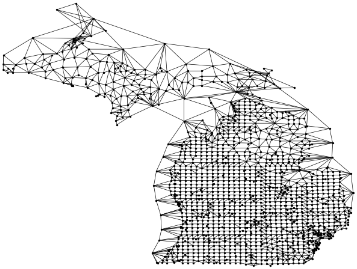
# Census Dual Graphs

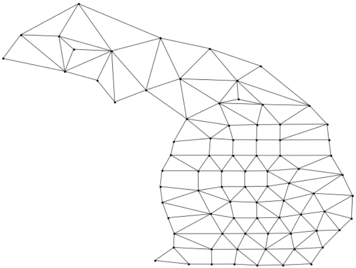




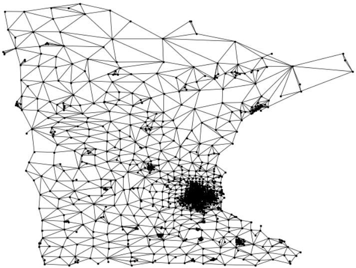


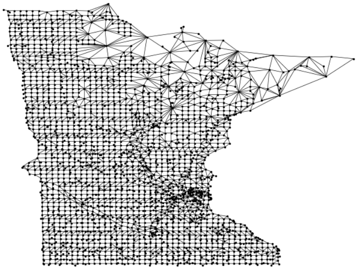
# Census Dual Graphs

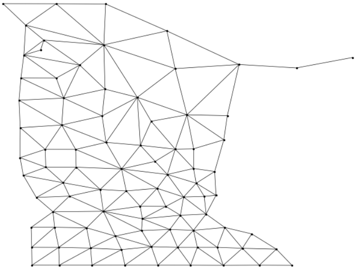




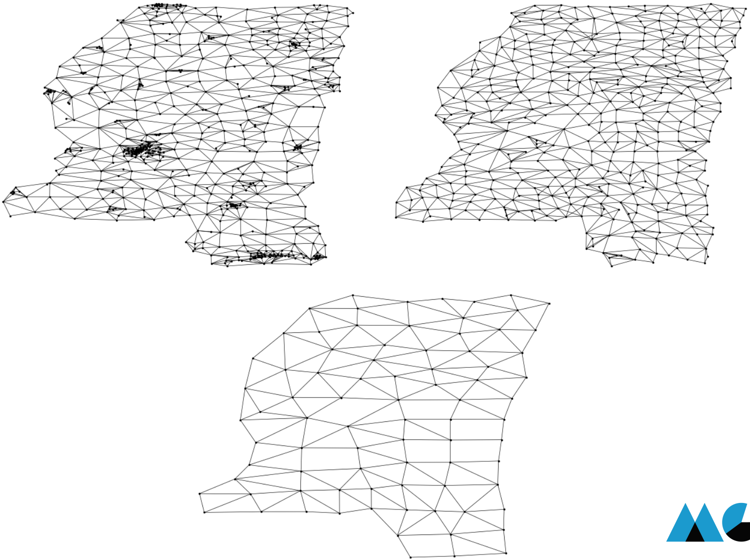
# Census Dual Graphs



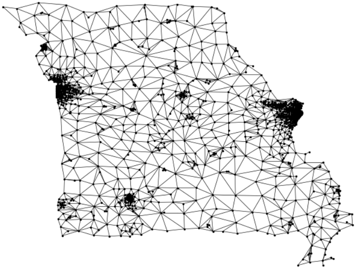


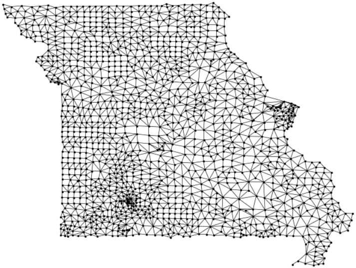


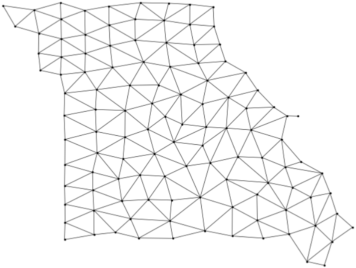
# Census Dual Graphs



# Census Dual Graphs







# Districting Plans

**Inputs:**

- A planar, connected graph $G = (V, E)$
- Weights $w : V \to \mathbb{R}^+$
- Population tolerance $\varepsilon$

**Output:**

- A partition $P = \{V_1, V_2, \ldots, V_k\}$ subject to the additional conditions:
    - $V_i \subset V$
    - $V_i \cap V_j = \emptyset$ for $i \neq j$
    - The induced subgraph of $G$ on $V_i$ is connected for all $i$ and
    - $(1-\varepsilon)\frac{\sum_{i=1}^{k}\sum_{v \in V_i} w(v)}{k} \leq |V_i| \leq (1+\varepsilon)\frac{\sum_{i=1}^{k}\sum_{v \in V_i} w(v)}{k}$

# Desirable Characteristics

**Example (What properties do we want?)**

# Desirable Characteristics

**Example (What properties do we want?)**

- Efficiency
- Parameter Variability
- Robustness
- Interpretability
- Mathematical Elegance
- All permissible plans are possible

# Legal "Requirements"

- Population Balance
- Contiguity
- Compactness
- Municipal Boundaries
- VRA Compliance
- Communities of Interest

# Legal “Requirements”

- **Population Balance**
- **Contiguity**
- Compactness
- Municipal Boundaries
- VRA Compliance
- Communities of Interest

# Legal “Requirements”

- Population Balance
- Contiguity
- **Compactness**
- Municipal Boundaries
- VRA Compliance
- Communities of Interest

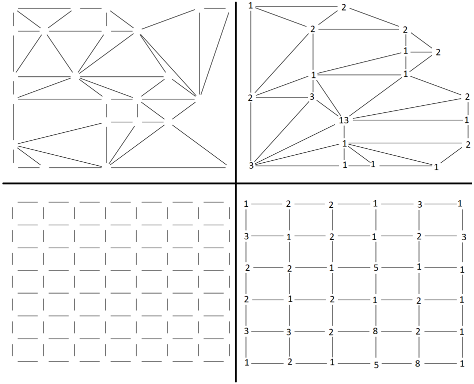
# Legal "Requirements"

- Population Balance
- Contiguity
- Compactness
- **Municipal Boundaries**
- **VRA Compliance**
- Communities of Interest

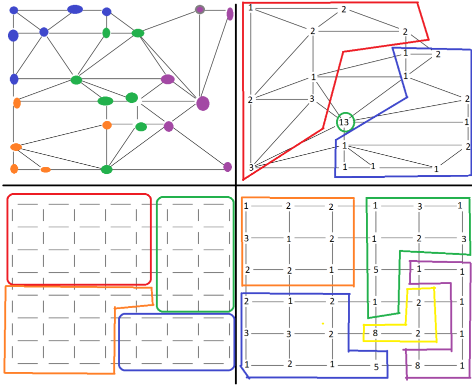
# Activity

Take a few minutes and partition the four graphs into the indicated number of districts. Think about how you might write an algorithm expressing your approach.

# Handout



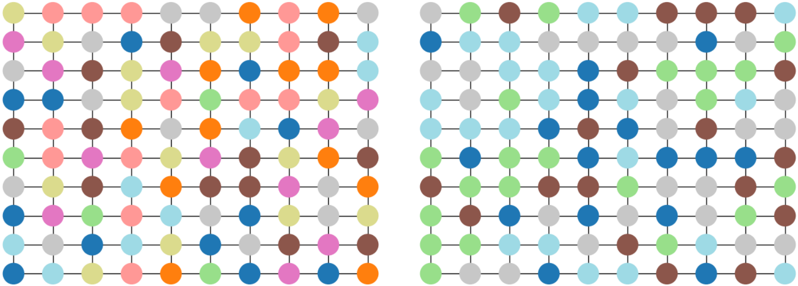
# Handout



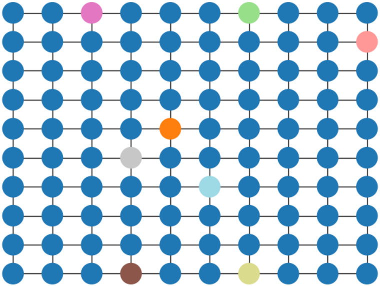
# MORAL:

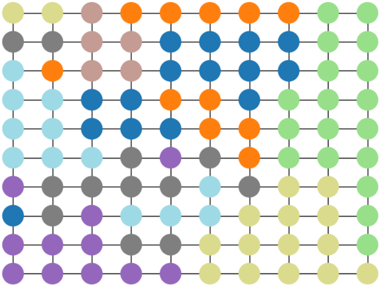
Computational Redistricting is NOT a solved problem!

# Toy Example: Random Assignment

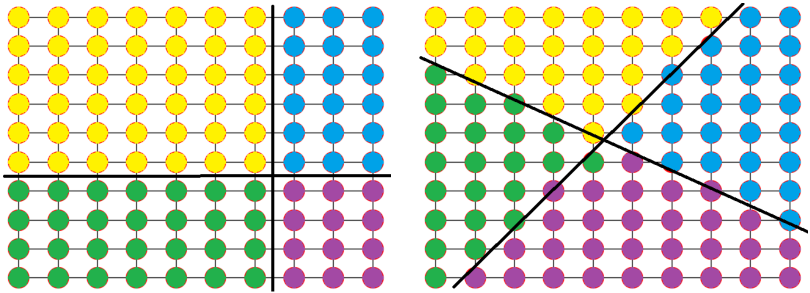


# Toy Example: Random Walkers

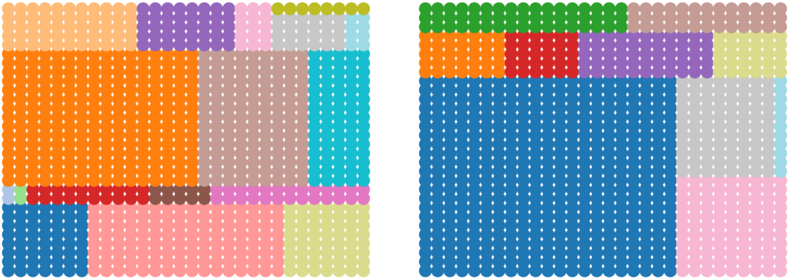




# Toy Example: Random Lines



# Toy Example: Random Rectangles



# Power Diagrams

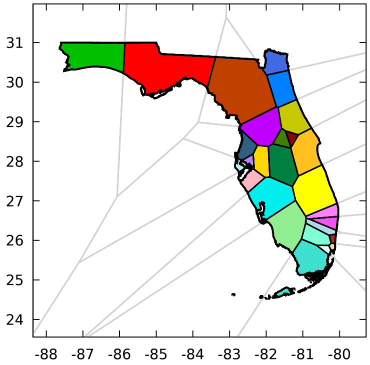



Figure 1: Florida (27 districts)

Figure: Power diagram for Florida: Balanced power diagrams for redistricting: V. Cohen–Addad, P. Klein, and N. Young.

# Other Straight Line Methods

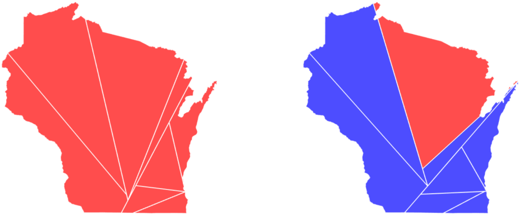


Figure: Split line partitioning of Wisconsin: Partisan gerrymandering with geographically compact districts: B. Alexeev and D. Mixon

- Split Line Methods
- Pretend that everything is a grid
- (Optimization) Draw lines even within households
- Alternatively, embed all voters on a circle

# Problems?

- No clean mapping on to discrete units
- Difficult to preserve municipalities, COI, VRA, etc.
- Assumes better control over data than actually exists
- Very hard to tune to arbitrary legal constraints

# Growing Districts

- Another popular class of methods are colloquially known as flood fills
- This procedure iteratively creates districts by growing them one node at a time
- Usually, contiguity is enforced at each step
- The process continues until the population is nearly balanced

# Flood Fill

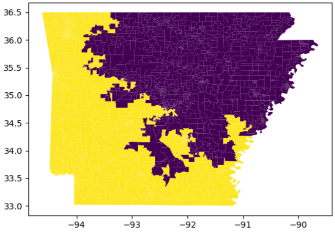
## Method

- *Select a node at random*
- *Select a random neighbor of the current cluster*
- *Alternatively, generate a list of neighbors and append sequentially*
- *Add if population allows and doesn't disconnect the complement*
- *Repeat until population balanced*

# Path Fill

## Method

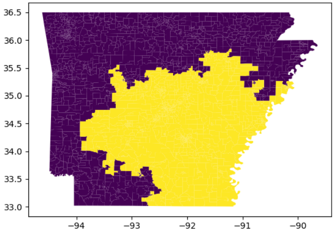
- *Start with an arbitrary node*
- *Select a node not in the district*
- *Add all the nodes on a shortest path from the new node to the district if it doesn't disconnect the complement or add too much to the population*
- *Repeat until population balanced*



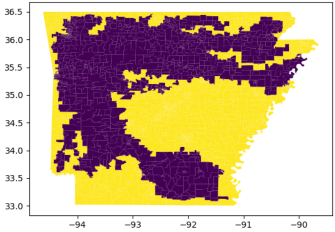
# Agglomerative

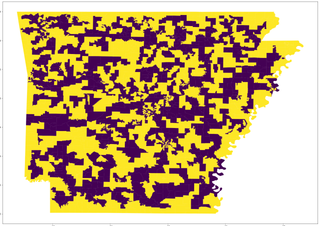
## Method

- *Start with each node in own component*
- *Select an arbitrary edge between two components*
    - *Merge clusters if population allows and doesn't disconnect the complement*
    - *If population doesn't allow, delete edge*
    - *If merging would disconnect the graph, merge the smallest population component*
- *Repeat until only 2 clusters*



# What can go wrong?

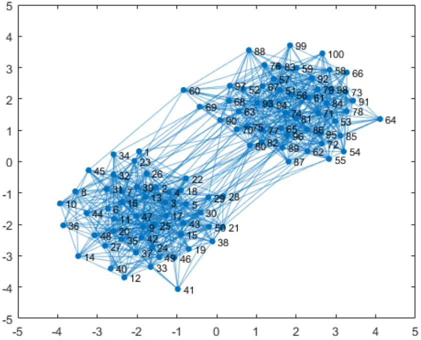


# Problems?

- High failure rate
- No control over distribution
- Medium hard to tune to arbitrary legal constraints
- Requires separate cleaning steps

# Network Clustering



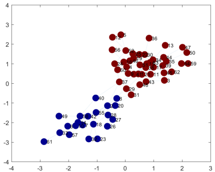
# What is a community?

# What is a community?

- Many intra–community links
- Few inter–community links
- Any measure that allows for dimension reduction
- Depth or closeness measures
- Different type of eyeball test

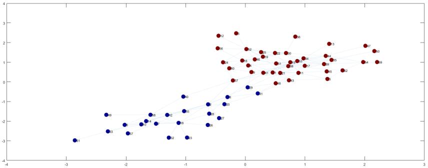
# Spectral Clustering

The idea behind spectral clustering is that communities should be sparsely connected to each other. This is usually defined in terms of an isoperimetric ratio, expressing the difference between the size of the boundary and the number of nodes in the community. The solution is given in terms of the eigenvectors of the Laplacian matrix.



# Modularity

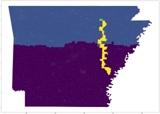
For modularity, we take the opposite definition. Now we define a community as a group of nodes that have more connections to each other than would be expected if we rewired the whole network. The solution is given in terms of the eigenvectors of the Modularity matrix.

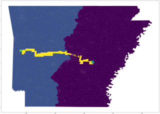


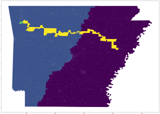
# Min Cut

## Method

- *Select random source and sink nodes*
- *Weight the edges in the graph by* $10^{min\ distance - 3}$
- *Compute the min cut*
- *Repeat until population balanced*



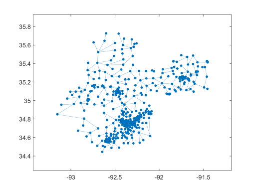


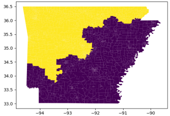


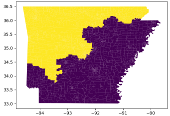
# Tree Partitions

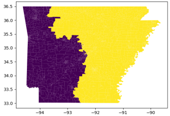
## Method

- *Generate a uniform spanning tree*
- *Cut an edge that leaves population balanced components*







# Problems?

- The underlying assumption for all of these methods is that the graph structure contains all of the relevant information for defining communities.
- However, for our setting, the useful information is usually annotations, not the nodes/edges themselves.
- For example, spectral clustering and modularity perform quite poorly on dual graphs that are very grid like
- Hard to optimize for many different functions at once

# Potential Solution

- Although the naive version of the network approaches seems poorly tuned for our setting there is some hope:
- These methods permit weighted generalizations that allow us to encode some measures of similarity between nodes
  - Demographics
  - Shared Geography
  - COI
  - Municipal Boundaries
- These weighted versions can then be interpreted as maximizing similarity within/minimizing similarity without and used to find larger partitions.
- Some success already, still a long way to go!

# Recursive Constructions

- Choose a methods for constructing a single (contiguous, population balanced, etc. ) a district
- Create one and repeat
- In general, bipartitioning, even in the unbalanced setting, is easier than k-partitioning
- Particularly true for many of the network methods, which tend to be significantly more stable for 2–partitions.

# Initial Seeds

- One use for these randomly drawn plans is as initial seeds for MCMC
- This provides a good heuristic check for convergence
- This can also solve data issues!

# MCMC

# The end!

Thanks!