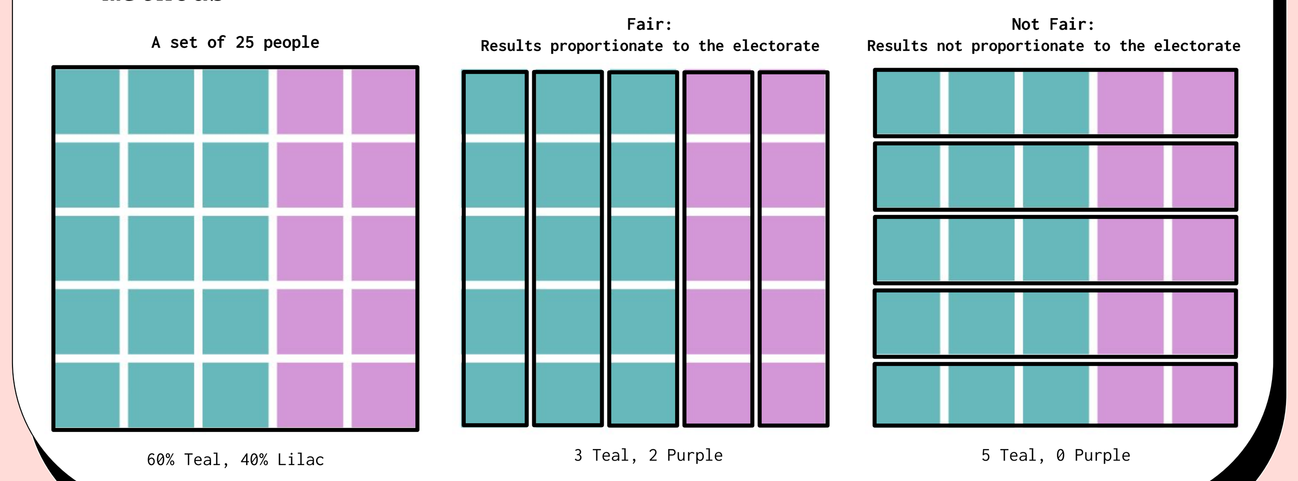
# CONNECTING THE DOTS

## An Exploration of Dual Graphs in Political Redistricting

Sarah Cannon, Mehrin Khan, Claire Vlases, Xinran (Joy) Zhang (Claremont McKenna College); Andy Emerson, Sara Anderson (Claremont Graduate University)

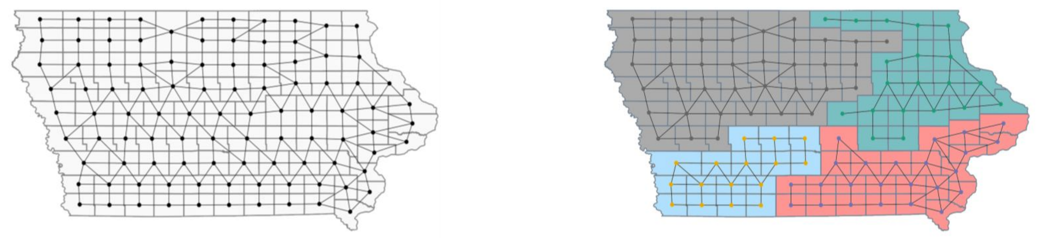
### Motivation

- Gerrymandering is when political districts are drawn in an **unfair way**
- Many mathematical tools exist for studying gerrymandering
- Nearly all **use *dual graphs* in their analysis**
- Little is known about the structure of dual graphs
- Learning more about dual graphs can **help improve methods** for studying redistricting and guide the development of new future methods




### Dual Graphs Background

- A way to **represent geographic** structure
- In Iowa: Make a vertex for each county, connect neighboring counties
- Can build redistricting plans out of this graph



For **more detailed** redistricting plans, use **finer levels** of geography

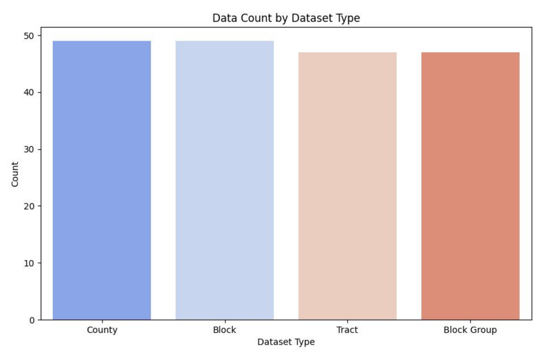
Counties | Census Tracts | Census Block Groups | Census Blocks

Bigger regions ⟷ Smaller regions

### Ethical Considerations

**Dual Graphs Considered:**

- From Redistricting Data Hub, prepared by Daryl DeFord (Washington State), which uses Census data
- Our data covers four levels of census geographies: counties, census tracts, census block groups, and census blocks. For each, we have data available from all states for which there were not data formatting issues:
  - Counties: All states except Nevada
  - Census tracts: All states except Nebraska, Nevada, and Wisconsin
  - Census block groups: All states except Nebraska, Nevada, Virginia, and Wisconsin
  - Census blocks: All states except Nevada




**Data validation:**

- Check total number of vertices is
- correct
- Check total population (added across all vertices) is correct

### Key Questions

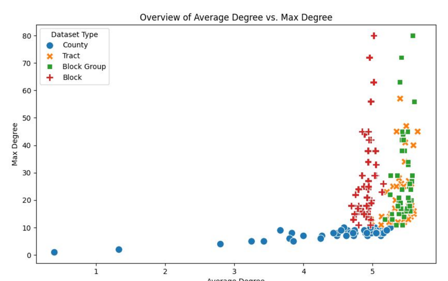
- How similar are **dual graphs to grids**?
- What **properties** do dual graphs have that are **similar** across all levels of geography?
- Are there any **key differences** between the different levels of geography?
- These graphs are sometimes described as "nearly triangulated and nearly planar." Is this true?

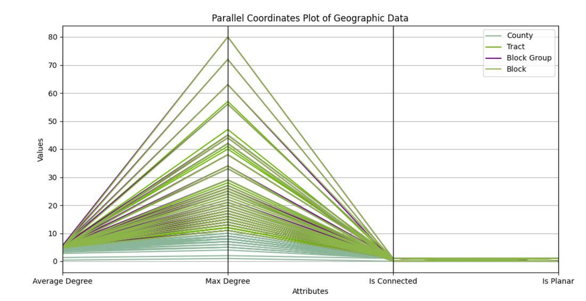
### Sequencing

A degree sequence is a list or sequence that represents the degrees of all nodes in a graph.

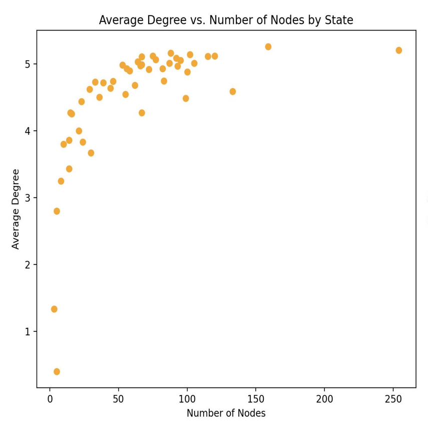
In a degree sequence, each element of the sequence corresponds to the degree of a node in the graph. The sequence is usually sorted in non-increasing order, meaning that the largest degree is listed first, followed by the second largest, and so on, until the smallest degree.

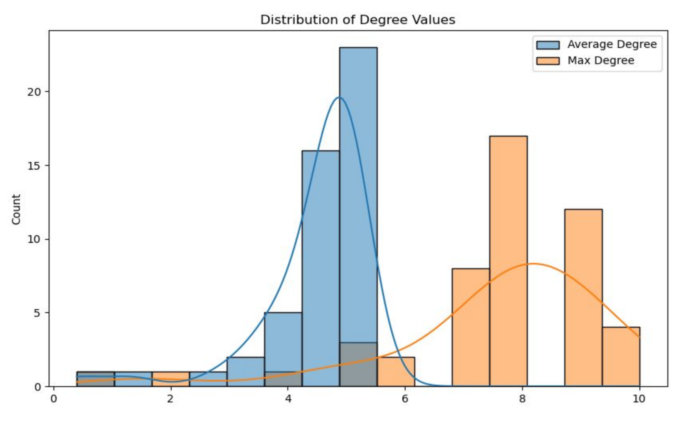
Looking at the entire data set: there are trends apparent in all 4 data sets, including a correlation between average and max degree.




Zooming in on county data: there are noticeable trends within the specific level too. For example, in county data, there is a correlation between average degree, max degree, and a distribution of degree values.



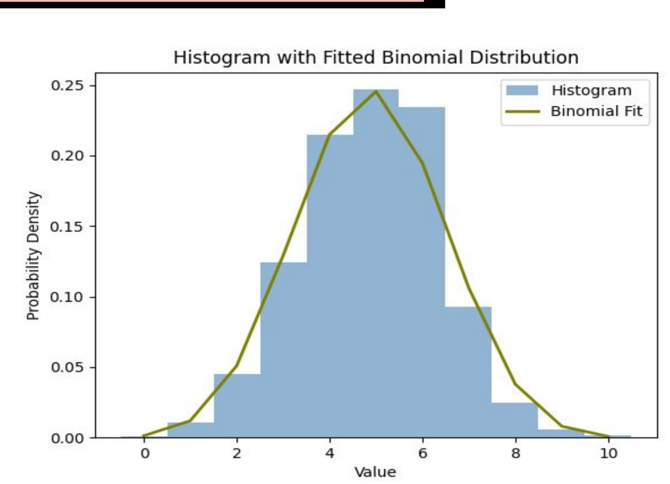




### Degree Distributions

**What is the distribution of degrees in a dual graph?**

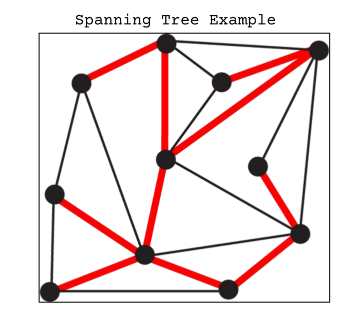
For county data, degrees appear to be **normally distributed**!

Notably, this pattern doesn't hold for other geographies due to the presence of large degree nodes.



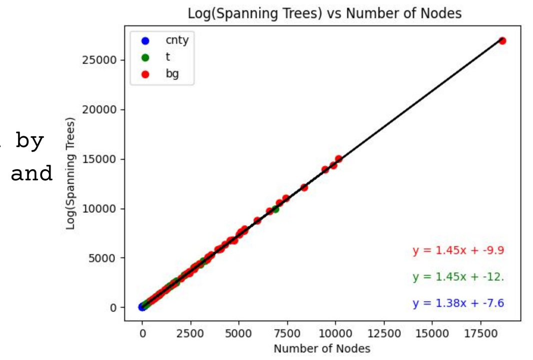

### Spanning Trees

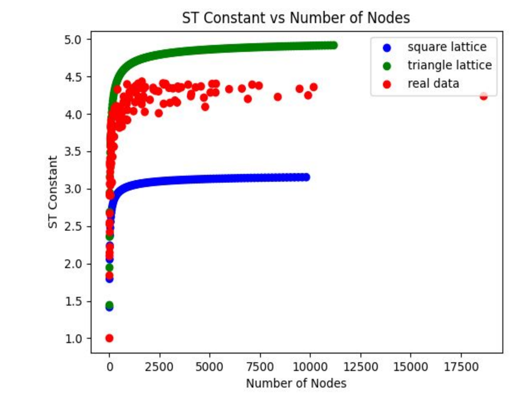
Many algorithms for studying redistricting rely heavily on spanning trees. **A spanning tree of a connected graph is a subgraph which uses all nodes and contains no cycles.** A disconnected graph will have no spanning trees, and a graph that is itself a tree will have exactly one spanning tree. **The more edges a graph has, compared to the number of nodes, the more spanning trees it will have.**



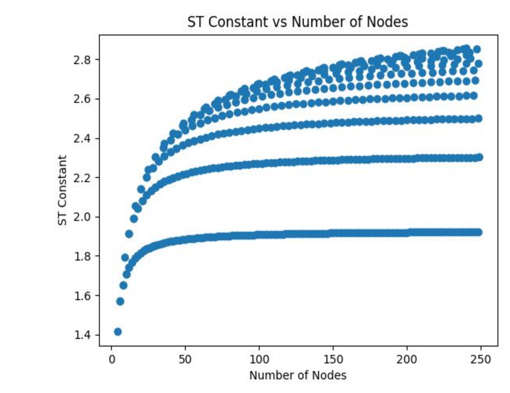

A large number of spanning trees indicates that a graph has many different paths between its vertices, and is well-connected.

We can use the number of spanning trees to calculate the **"spanning tree constant,"** given by $S^{1/n}$ where S is the number of spanning trees and n is the number of nodes in the graph. For a square lattice, as we increase the size, our spanning tree constant converges somewhere between 4 and 4.25. For a triangular lattice, this number converges to just under 5.




Our data falls between a square grid and a triangular grid

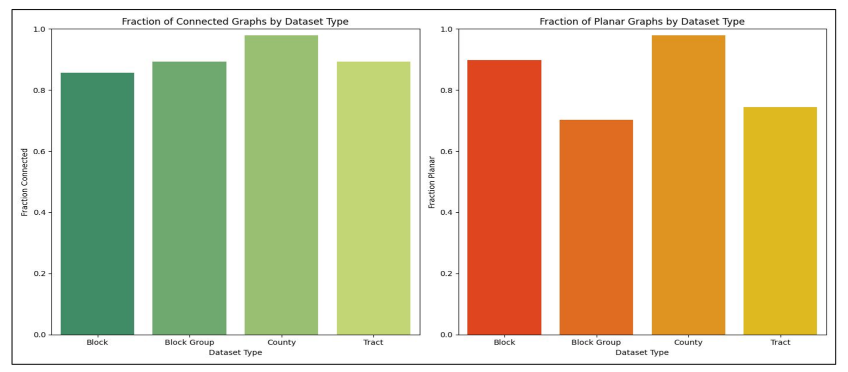


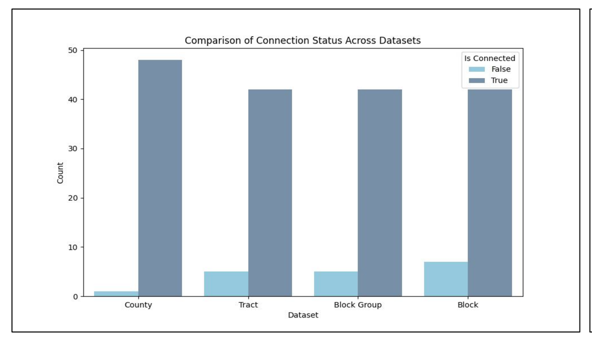

The shape of the grid has a large effect on the spanning tree constant

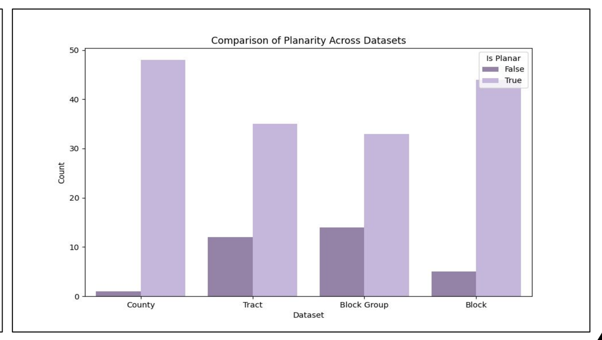
### Connectivity and Planarity

A graph **is connected if for every pair of nodes there exists a path between them.** A graph **is planar if it can be drawn in the plane such that no edges cross.** While one might expect dual graphs to be both connected and planar because they arise from real-world geographies, there are many reasons why they might not be.

For example, a graph might not be connected because a state has islands; a graph might not be planar because five geographic regions meet at a single point, resulting in a complete graph $K_5$ as a subgraph of the dual graph.



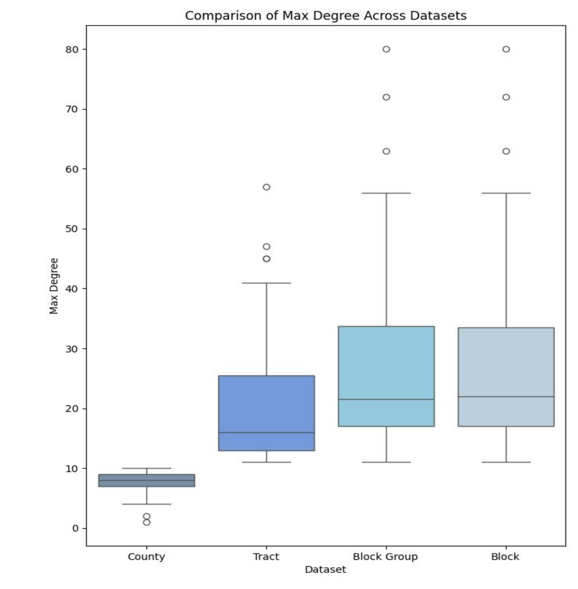



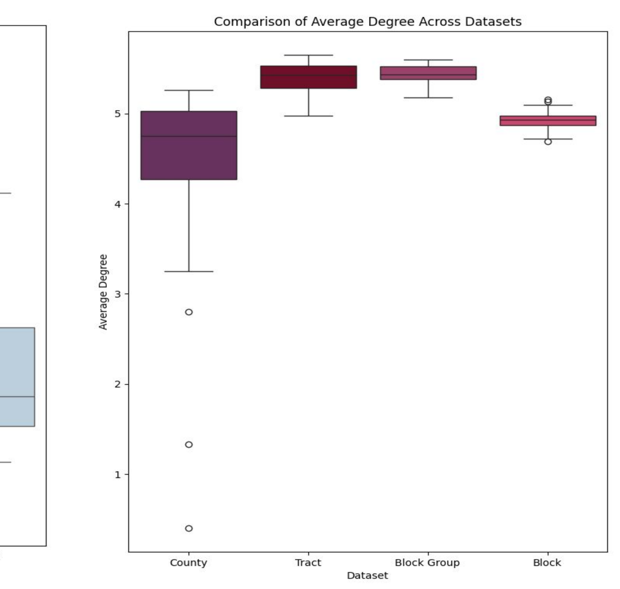




There is a variation of connectivity and planarity across the four datasets

### Degrees

The **degree of a node is the number of edges incident to it,** which is equivalent to the number of connections or neighbors that the node has. The degree of nodes is one of the most fundamental measures in the study of a graph's underlying structure.

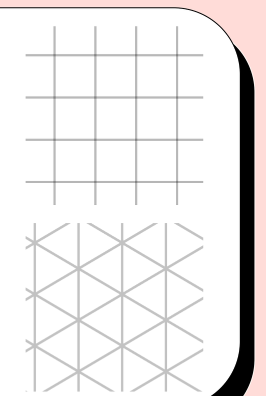



- We expected average degree to increase as graphs got larger, so we were **surprised to see block graphs had smaller average degree.**
- Square grids have max degree 4 and average degree just less than 4; triangular grids have max degree 6 and average degree just less than 6
- In terms of degrees, **our graphs are closer to triangular grids, but have significantly larger max degrees**

### Key Takeaways

- We understand the **typical statistics** of dual graphs
- County data has different features than the other levels of geography
- Block data has some different features than the other levels of geography
- Dual graphs fall somewhere **between square grids and triangular grids** in terms of number of spanning trees
- The term **"nearly triangulated"** may be apt; the term **"nearly planar"** is clearly not: a significant **majority of dual graphs are planar!**



### Next Steps

**Splitability of spanning trees**: A new polynomial-time algorithm for sampling districting plans on grids relies on the key fact that a polynomial fraction of spanning trees on grids can be split exactly in half. For our real-world graphs, how likely is it that a random spanning tree can be split exactly in half?

**Clustering coefficients:**

- Local: The local clustering coefficient measures the probability that neighbors of a given node are also connected to each other, forming a triangle. For a node i, the local clustering coefficient $C_i$ is defined as:

$$C_i = \frac{2e_i}{k_i(k_i-1)}$$

  where $e_i$ is the number of edges between the neighbors of node i, and $k_i$ is the degree of node i (the number of neighbors). High local clustering coefficient for a city would indicate that its neighboring cities are also well connected among themselves, forming a tightly knit regional network. This could be important for logistics planning, for example.
- Global: This could be indicative of strong regional cohesion or the presence of natural barriers that limit inter-regional connections.

**Centrality measures** are used to identify the most important vertices within a graph with respect to different utilities.
These measures provide a way to quantify the significance of individual nodes based on their position in the network. We want to continue by exploring the four most common types of centrality: degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality.