Trees project

VRDI

July 29, 2019

Spanning-tree-based methods for partitioning graphs into districts

Setup

We have a graph G in which we would like in k districts, a total population P giving ideal population $\frac{P}{k} = i$, and a tolerance ε for population deviation. Let $p(v)$ be the population of a vertex $v \in G$ and $p(H)$ be the population of a sub-graph $H \subseteq G$.

Method π

1. Pick a spanning tree T of G uniformly at random.
2. Find an edge of T such that the population is split evenly into two, or as close as possible, by its removal. Call this the center c. Choose a random leaf ℓ, with population $p(\ell)$. Let this be our initial sub-graph $H = \{\ell\}$.
3. From the starting leaf ℓ, walk one node in towards the center. Update H to include the edge walked along, the node stopped at (ℓ'), and all parts of the graph downstream from ℓ'. Check $p(H)$.
4. While $p(H) < i - \varepsilon$, repeat.
5. As soon as $p(H) \geq i - \varepsilon$, stop changing H. Walk one edge closer to the center, and call the new sub-graph generated by this step H'. If $|p(H') - i| \geq |p(H) - i|$, stop and revert to H. Cut H off from the larger spanning tree, and use it as the first district.
6. If instead $p(H')$ is closer to i than $p(H)$, throw H out and consider H' instead. Repeat the above process; walk one step closer to the center, and check if the population of this new sub-graph H" is closer to or farther from the ideal population. Repeat until we've gotten a a sub-graph with population as close to i as possible, given the randomly chosen starting leaf. Use this sub-graph for our first district.

Once our first district is determined, we have two options to finish constructing the districting plan.

Method π.1

Draw a new, random spanning tree T' for $G \setminus H$, where H is the first district as generated above. Repeat the above process on the new tree, drawing a new tree after cutting off each subsequent district.

Method π.2

Proceed with the original spanning tree T of G, cutting off each new district as it is generated through the process above, but otherwise leaving it untouched.

Method 3

1. Pick a spanning tree T of G uniformly at random.
2. **While** no leaf has population in $[i-\varepsilon, i+\varepsilon]$:
   - For each leaf ℓ of T (traversed in random order, let's say), retract ℓ to its parent (unique neighbor in T) $\text{pred}(\ell)$ and add its population to its parent's, so that $p_{\text{new}}(\text{pred}(\ell)) = p(\text{pred}(\ell)) + p(\ell)$
3. Pick a random leaf ℓ with $p(\ell) \in [i-\varepsilon, i+\varepsilon]$ and cut it off, making it and the vertices that it has absorbed into a new district. Repeat steps 1 and 2 for the subgraph that remains, unless there remains only one more district to be formed, in which case define the final district to be the remaining subgraph.

Method 4

1. Pick a spanning tree T of G uniformly at random. Initialize an empty set $L = \{\}$ of node labels.
2. **While** there remain multiple districts to form:
   - For each leaf ℓ of T:
     - Retract ℓ to its parent (unique neighbor in T) $\text{pred}(\ell)$ and add its population to its parent's, so that $p_{\text{new}}(\text{pred}(\ell)) = p(\text{pred}(\ell)) + p(\ell)$
     - If $\ell \in L$, then check whether $|p_{\text{new}}(\text{pred}(\ell)) - i| < |p(\ell) - i|$ and check whether $\text{pred}(\ell)$ becomes a leaf by this retraction (i.e. has no other children). If both conditions are satisfied then accept the retraction, otherwise undo it and cut off ℓ, making it and the vertices that it has absorbed into a new district.
     - If $p_{\text{new}}(\text{pred}(\ell)) \in [i-\varepsilon, i+\varepsilon]$, then add $\text{pred}(\ell)$ to L.
     - If $p_{\text{new}}(\text{pred}(\ell)) \geq i+\varepsilon$, then throw the tree out and start the process again.
3. When there is only one district left to form, let the remainder of the existing tree be that district.

Proof of non-subset-separability for 4×2 grid with only one spanning tree

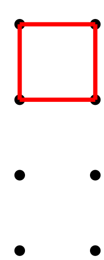
We know from the work of the Networks week 5 project group that we can uniquely partition any spanning tree on this graph into 4 equal parts.

Definition 1. *Let G be a finite graph, k be a positive integer, and $\mathcal{P}_k(G)$ be the set of partitions of G into k connected subgraphs, called "districts." If $\ell < k$ and H is a subgraph of G that can be formed as the the union of ℓ districts for some valid districting plan in $\mathcal{P}_k(G)$ then let $\mathcal{P}_{k,\ell}(G,H) \subset \mathcal{P}_k(G)$ be the set of partitions of G in which H is exactly the union of ℓ districts in the partition. A probability measure π on $\mathcal{P}_k(G)$ is* subset-separable *if for any $\ell < k$ and subgraph H that is the union of ℓ districts, then $\pi|_{\mathcal{P}_{k,\ell}(G,H)}$ induces independent marginal distributions on $\mathcal{P}_\ell(H)$ and $\mathcal{P}_{k-\ell}(G \smallsetminus H)$ (where $\pi|_{\mathcal{P}_{k,\ell}(G,H)}(S) := \pi(S)/\pi(\mathcal{P}_{k,\ell}(G,H))$ for a measurable set $S \subset \mathcal{P}_{k,\ell}(G,H)$). In other words, if $d \in \mathcal{P}_{k,\ell}(G,H)$ is a districting plan and $d|_H \in \mathcal{P}_\ell(H)$ and $d|_{G \smallsetminus H}\mathcal{P}_{k-\ell}(G \smallsetminus H)$ are its restrictions to the subgraph and its complement, then $\pi|_{\mathcal{P}_{k,\ell}(G,H)} = \pi_H(d|_H)\pi_{G \smallsetminus H}(d|_{G \smallsetminus H})$ where π_H and $\pi_{G \smallsetminus H}$ are the independent marginal distributions on $\mathcal{P}_\ell(H)$ and $\mathcal{P}_{k-\ell}(G \smallsetminus H)$, respectively.*

Lemma 1. *For any tree (or forest), there is at most one way of partitioning it into districts (connected subgraphs) of size 2.*

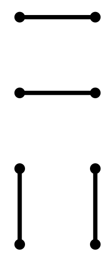
Proof. A partition into size-2 districts is equivalent to a perfect matching. There is an easy proof that a tree admits at most one perfect matching based on taking off leaves and inducting on the number of vertices; see this Stack Exchange post or the third problem on this midterm. ☐

We claim that no spanning tree method of partitioning the graph G with fixed sub-graph H, shown in red below, is subset-separable.

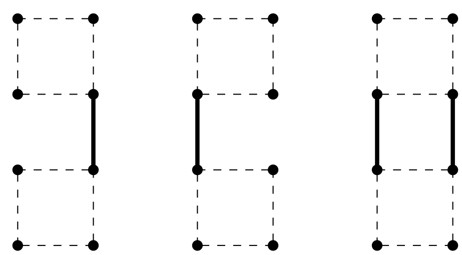


Proof. We'll begin by enumerating the spanning trees for G which allow partition into 4 districts, exactly 2 of which are in H. Consider that each of H and $G \setminus H$ can be partitioned into either two vertical districts or two horizontal districts, since the districts must be contiguous. See figure below for the two options.

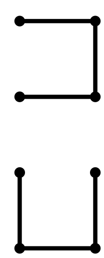
In order to produce an allowable districting plan in H and $G \setminus H$, any spanning tree for G must have as two sub-graphs the spanning trees of H and



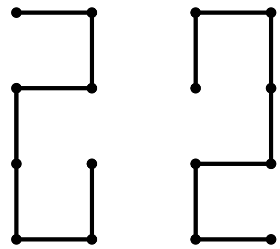
$G \setminus H$. There are three possible ways we can connect these two sub-graphs: with one vertical edge on the right, one on the left, or both, as shown below.



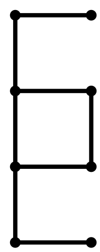
We'll break into cases based on whether H and $G \setminus H$ are connected by one or two edges. Consider that a spanning tree on a square of four nodes looks like a "U"; see figure below for some possibilities.



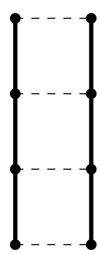
In the case with only one vertical connector, every rotation of the spanning tree of H is compatible (i.e. can be cut to produce an allowable districting plan) with every rotation of the spanning tree of $G \setminus H$. So there are $32 = 4 \times 4 \times 2$ ways of drawing a spanning tree with only one vertical connector between the two sub-graphs that works: 4 possible rotations on top, times 4 on the bottom, times 2 choices for which side to place the connecting edge on. See below for a sample of possible trees.



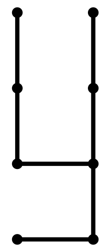
Now we'll look at the case with two vertical connectors. Since both vertical sides are present on the middle square, it's not possible for both spanning trees of H and $G \setminus H$ to be oriented horizontally, as that would produce a loop. See below.



So we'll break this into more cases based on whether the districting plans for H and $G \setminus H$ are both vertical. If this is true, then we know the edges shown below must be in the tree. It's then only a question of where to place the single horizontal edge that is necessary to make the graph a spanning tree. There are exactly 4 choices for this; see the dashed edges below.



If one is vertical and the other is horizontal, then we have two choices of which way to orient the horizontal "U", and two choices of where the horizontal districts live: in H or in $G \setminus H$. This gives a total of 4 more spanning trees that are compatible with the required districting plan.

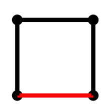


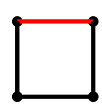
Thus, there are 40 ways to draw a spanning tree of this eight-node graph that can be partitioned into four allowable districts. Let's now consider the conditional probability of having a pair of either vertical or horizontal districts in H and $G \setminus H$. Table 1 shows the number of ways to get vertical/horizontal districts in H in the columns, and on $G \setminus H$ in the rows.

	plans on H		
	vertical	horizontal	
vertical	12	10	22
horizontal	10	8	18
	22	18	40

From the perfectly formatted table above, we can calculate the probability that $G \setminus H$ will have two horizontal districts, given the districting plan of H. If H is cut into two verticals, $G \setminus H$ has a $10/22$ chance of being paritioned into horizontal districts. If H is instead horizontal, $G \setminus H$ has a $8/18$ chance of being horizontal. Since $10/22 \neq 8/18$, we can see the probability distributions on H and $G \setminus H$ are not independent. Therefore, any method that uses a single spanning tree to partition G into districts is not subset-separable.

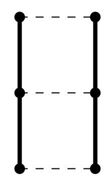
What if instead we use a randomly generated (allowable) spanning tree to draw the first district, then throw the tree out and draw another after each step? This method is not subset-separable either. Using this method, there are three ways to draw an allowable first district in one of H or $G \setminus H$ in this case: it must either be one of the vertical districts, or an outer horizontal district. The options drawn in red below are not allowed, since they would leave the remainder of the spanning tree disconnected.



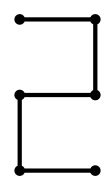


Let's consider what happens if our first district drawn is an outer horizontal (wlog, the bottom-most horizontal). We'll consider all possible spanning trees on the 3×2 grid that remains when we cut off the bottom two nodes.

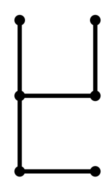
Either both top and bottom have two vertical edges, two horizontals, or one of each. If there are two sets of vertical edges, we have three ways of making them into one spanning tree: one horizontal in each possible place, see below.



if two horizontals, we have two options for which way each "U" is facing, which makes 4 different spanning trees total.



If one of each, there's two options for U-orientation if the pair of horizontal edges is on top, and two for if it's on the bottom, which makes 4 other spanning trees, for 11 total.



Out of these 11, only the ones which already include the bottom-most horizontal edge are actually allowable, since we already know the districting plan for $G \setminus H$ is a pair of horizontal districts. These leaves us with 7 options for the top district, 3 of which are the vertical plans and 4 of which are horizontal. Thus the top districting plan has a probability distribution dependent on the bottom districting plan, so this method is not subset-separable either.

$\square$