**Application of a famous mathematical theorem to the question of what is the minimum number of county splits in redistricting maps**

The number of splits is defined by summing the splits in each county where a county has n – 1 splits if it has pieces of n districts. Splits.docx argues that this is the best simple metric.

One could argue that the question is ill-posed because there are simple examples that have what I will call accidental degeneracy.  The simplest example is a county that has exactly the right number of people for a CD. Accidental degeneracy also occurs if a number n of counties has exactly the correct number of people for q districts. This has not been found to happen in PA, although there are a few near degeneracies where there only a few thousand too few or too many people.  Most of us would say that is good enough, but the courts have ruled that even a deviation exceeding one is unacceptable and this criterion has been applied rigorously in PA, even to the extent of splitting precinct and census blocks.  While this is really an insane criterion, it makes it easier to derive a robust answer for Sm because the probability of accidental degeneracy is so small with districts having of the order of 700,000 people, depending upon the state.  While enough accidental degeneracy could reduce Sm, even down to 0, it turns out that Sm has a well-defined value if there is no accidental degeneracy.

Let us define a graph for a districting plan.  Each vertex in the graph represents a district.  Each edge connecting two districts represents a county that is split between those two districts. A county that is split s times will have s associated edges. There is latitude in drawing the edges for a county that is split more than once, but all drawings are equivalent for obtaining Sm. The fundamental graph theorem, traceable back to Euler, is written with the usual symbols for number of edges e and number of vertices v, where c is the number of cycles and p is the number of disjoint pieces.

                                    e – v = c – p                            (1)

With no accidental degeneracy, the graph is connected and p = 1.  A cycle occurs when there is more than one path of edges between two vertices.  Rewriting Eq. (1) in terms of the number of districts D=v and number of splits S=e, one has

                                    S = c + D -1                            (2)

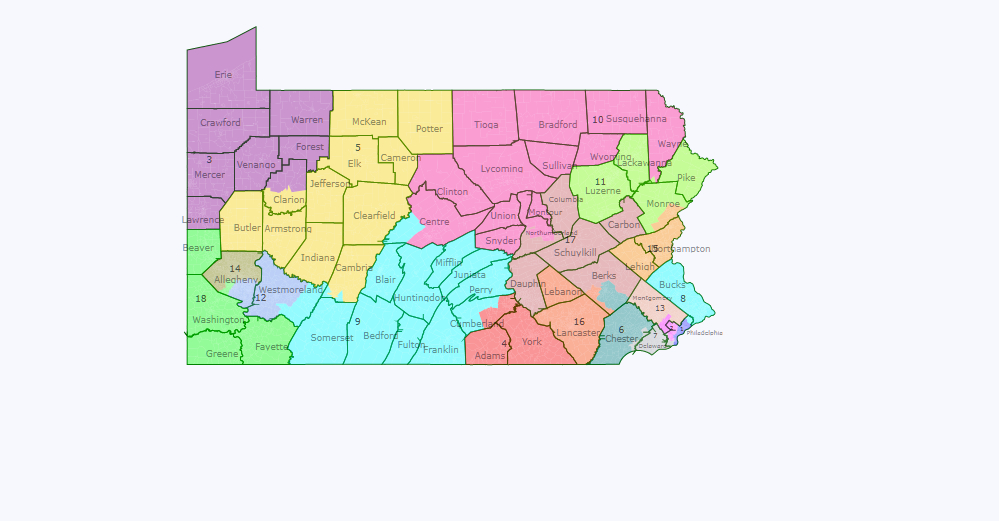
S in Eq. (2) is minimized when c = 0.  (Incidentally, such graphs are called trees in graph theory.)  This gives a simple answer to our original question

                                   Sm = D -1                            (3)

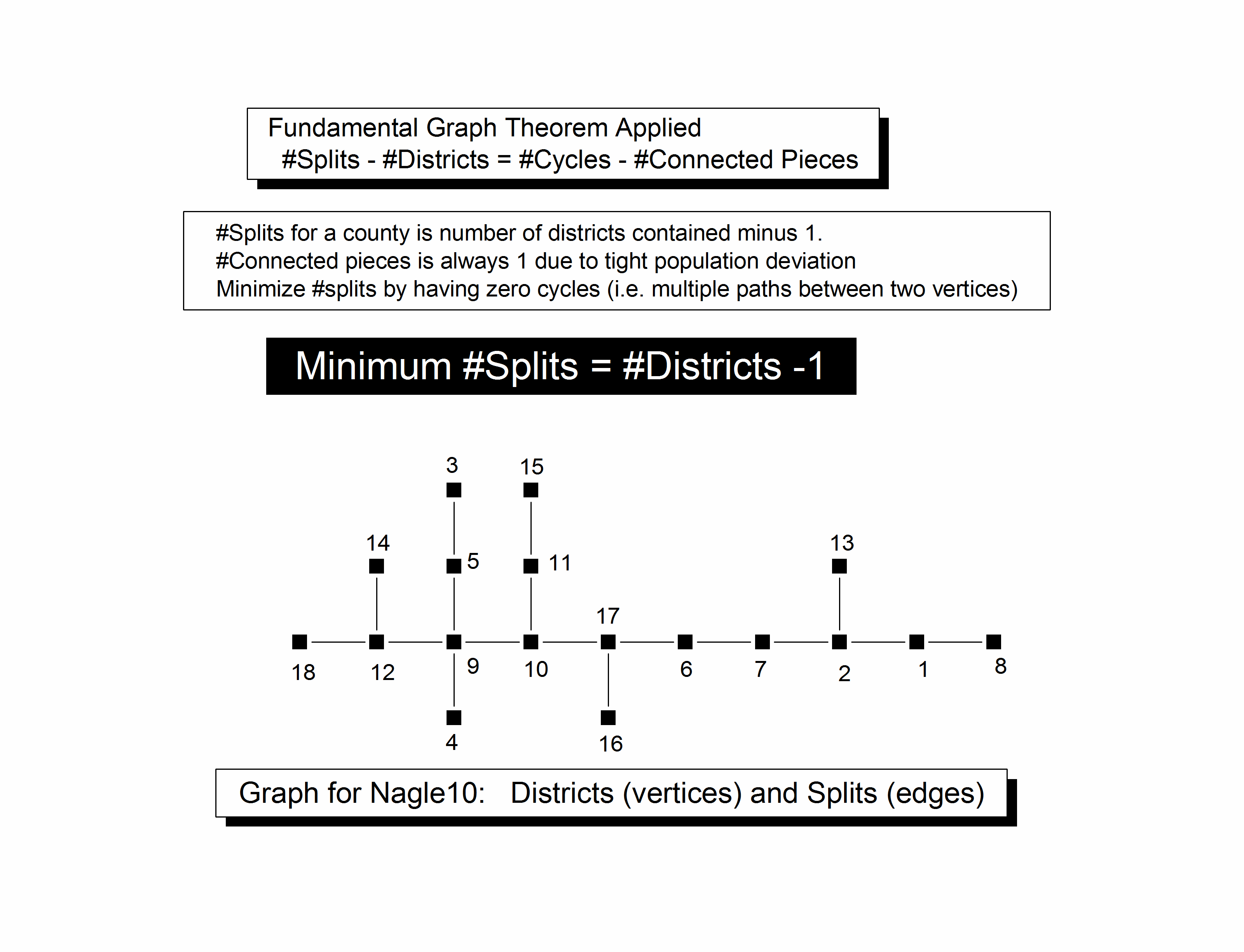
**Question: Is Eq. (3) well-known among districters?  If so, please send me a reference.** It does seem that those dedicated to this criterion have intuited it.

It is also an interesting question whether Sm can always be achieved. I am grateful to Fred Murphy for informing me that the minimum can always be achieved.

Start by drawing an arbitrary plan. If it contains cycles, choose any cycle c’ and proceed to break it while not adding more cycles elsewhere as follows. Let the number of districts in c’ be C, which is also the number of splits in c’. Starting with any district in c’, assign indices j to the districts and to the splits sequentially around c’ such that the districts (vertices) j and j+1, modulo C, are joined by split j. Determine the population of each of the pieces in each split county, p(j,j) and p(j+1,j). Find the minimum pmin in the set {(p(j,j),p(j+1,j),j=1,…,C modulo C}. If pmin corresponds to p(j,j), then shift all of pmin= p(j,j) in county j and district j to district j+1. This unsplits county j but it temporarily decreases the population of district j and increases the population of district j+1. To equalize all the populations of all the districts in c’, follow with subsequent shifts of population pmin from p(j+m,j+m) to p(j+m+1,j+m), m=2,..,C. Each shift is possible because from pmin < p(j+m,j+m), m=2,…,C. Barring accidental degeneracy of pmin, only the split of county j is removed. Finally, if pmin corresponds to p(j,j+1) instead of p(j,j) the same coordinated shifts in population proceed in the opposite direction around c’. Then, one proceeds to any remaining cycles c” in whatever order one likes.

Following is an example of a map with S = 17 that I have drawn for PA which has D = 18.

Following is the graph for this map.



This is a simple, clean result for minimum splits. However, if it is decided that larger population deviations can be allowed in order to better satisfy other criteria, then it becomes more probable that near degeneracies can be found. A favorite pastime of some map drawers is to search out groups of contiguous counties that have populations close to an integral number q of districts. Each such group then breaks the tree into another piece, so p increases by one and the number of splits decreases by one. I have seen such maps that have four such pieces, thereby claiming to have achieved only 14 splits in PA instead of 17. Of course, this tends to decrease the compactness of the districts, and it requires larger population deviations.

It is, nevertheless, an amusing operations research problem to design an algorithm to search out near degeneracies given an allowable population deviation. It is then an interesting mathematical problem to determine whether the algorithm is efficient. One might begin by assigning random populations to abstract counties in a state with population P, Q counties and D districts and then test whether any combination of q counties has population nP/q ± n appropriate for n districts with an average county deviation less than . However, most solutions would be invalid because the counties would not likely be contiguous. More subtly, even contiguous counties could split the map such that an acceptable deviation could not be achieved in the remaining parts; the simplest example is if the putative combination spatially isolates a small county at the edge of the map. One might then address a spatial map of a real state numerically, or even arbitrary planar maps with only triangular faces in the interior.

Finally, splits.doc mentions metric #7: Minimize the sum of populations in split counties but don’t add the population of that part of split counties that has the most people. Although arduous, this seems like a reasonable metric. However, it rewards finding groups of counties with near degeneracy because then achieving zero population deviation only adds a small number to the sum. But, as noted at the end of the penultimate paragraph above, this will tend to decrease compactness.