Chapter Four

The set P

Squeezing more information from the set of optimal solutions

Paul Anderson and Amy N. Langville

The set P is the set of all optimal solutions to the rankability integer program of Chapter 3. In this chapter, we squeeze as much rankability information as we can from this set.

4.1 VISUALIZING THE SET P

As the examples of Chapter 3 showed, the set P can be large. Even for small sets with, say $p = 10$, the best way to display all 10 optimal rankings is not obvious. This section introduces a **Markov chain** to visualize aspects of the set P.

The set P contains $p = |P|$ rankings of n items. In this section, we "think graphically" about P. First, we create a graph G with $n + 2$ nodes, one for each of the n items plus two dummy nodes, a sink node labeled 0 and a source node labeled $n+1$. Next we add directed links between nodes as follows. A link g_{ij} from node i to node j is the proportion of rankings in P that have node i directly above node j in rankings in P.

Let's use an example to make these ideas precise. For the $n = 4$ input matrix of dominance relations,

$$\boldsymbol{D} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix} \end{array},$$

the algorithm of Chapter 3 produces optimal rankability information $k = 4$,

$p = 6$, and

$$P = \left\{ \begin{pmatrix} 3 \\ 4 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 4 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 4 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 2 \\ 3 \\ 1 \end{pmatrix} \right\}.$$

Given this information, we can then create a matrix $\boldsymbol{G}$ representing the directed graph.

$$\boldsymbol{G} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 0 & 1/3 & 1/2 & 1/6 & 0 \\ 0 & 0 & 1/6 & 0 & 1/3 & 1/2 \\ 0 & 1/6 & 0 & 1/2 & 0 & 1/3 \\ 0 & 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 1/3 & 1/2 & 0 & 0 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

The $(1, 2)$-element $g_{12} = 1/6$ because only one of the six rankings in P has item 1 ranked directly above item 2. In addition to its matrix representation, graph $\boldsymbol{G}$ can also be displayed pictorially with the source node 0 above all nodes and the sink node $n + 1$ below. See Figure **??**.

TO DO. Maybe Paul knows software to make pretty graph FIGURE

The sink node has $g_{02} = 1/3$ because two of the six, so one-third, of the rankings in P have item 2 in the first rank position. Similarly, source node 5 contains information about the last place position in the ranking. Thus, $g_{25} = 1/3$ because two of the six rankings have item 2 in the last position.

Forming $\boldsymbol{G}$ in this way produces a Markov chain because each row is a stochastic vector of probabilities that sum to 1. State $n + 1$ is an absorbing state since all flow ends up there eventually. Notice that each ranking in P is a path of length $n + 1$ through $\boldsymbol{G}$ from 0 to $n + 1$. One common measure for a Markov chain with an absorbing state (or states) is **time to absorption**. One can compute the expected number of steps to absorption given that the chain starts in a given transient state. We apply this to the above example $\boldsymbol{G}$ and rank the states by their time to absorption.

$$\boldsymbol{G} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ \hline 5 \end{array} \left(\begin{array}{c} \begin{array}{ccccc|c} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{array}{c} \\ 0 \\ 0 \\ 0 \\ 0 \\ \hline 0 \end{array} \begin{array}{c} \begin{array}{ccccc|c} 0 & 1/3 & 1/2 & 1/6 & 0 \end{array} \\ \begin{array}{ccccc|c} 0 & 1/6 & 0 & 1/3 & 1/2 \end{array} \\ \begin{array}{ccccc|c} 0 & 1/6 & 0 & 1/2 & 1/3 \end{array} \\ \begin{array}{ccccc|c} 0 & 1/2 & 0 & 1/2 & 0 \end{array} \\ \begin{array}{ccccc|c} 0 & 1/3 & 1/2 & 0 & 1/6 \end{array} \\ \hline \begin{array}{ccccc|c} 0 & 0 & 0 & 0 & 1 \end{array} \end{array} \end{array}\right) = \left(\begin{array}{c|c} \boldsymbol{T}_{11} & \boldsymbol{T}_{12} \\ \hline \boldsymbol{T}_{21} & \boldsymbol{T}_{22} \end{array}\right),$$

where $\boldsymbol{T}_{11}$ is the submatrix corresponding to the transient states 0 through 4 and $\boldsymbol{T}_{22}$ is the submatrix, in this case, scalar, corresponding to absorbing state

5. The vector $\boldsymbol{t} = (\boldsymbol{I} - \boldsymbol{T}_{11})^{-1}\boldsymbol{e}$, where $\boldsymbol{e}$ is the vector of all ones, contains the expected number of steps to absorption given that the chain starts in a particular transient state. For this example,

$$\boldsymbol{t} = \begin{pmatrix} 5 \\ 2.86 \\ 3.64 \\ 4.32 \\ 3.77 \end{pmatrix}$$

so that the ranking of nodes from first to last by time to absorption is [3 4 2 1], which is a member of P.

Ranking the nodes by the time to absorption vector $\boldsymbol{t}$ makes sense since a node that has a short time to absorption must be in or near last place in most of the rankings in P, whereas a node with a long time to absorption must be very far from last place, and hence, near first place. *The ranking given by $\boldsymbol{t}$ can be considered an aggregate of the rankings in P.* Symmetrically ordering $\boldsymbol{G}$ by this ranking produces, for rankable data, a banded matrix whose spy plot shows most of the mass near the diagonal. Ordering the matrix this way makes it clear where, for less rankable data, the indecision in P occurs. See Figure 4.1.

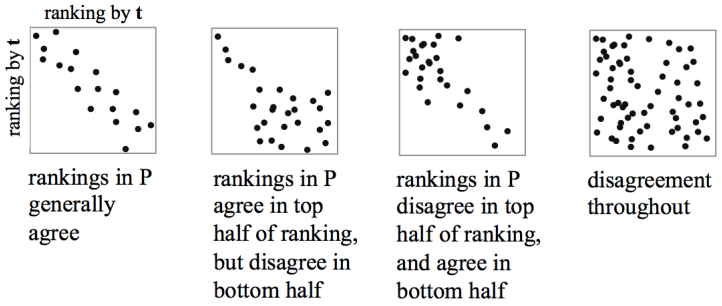

Figure 4.1: **Visualizing the set P.** Symmetrically reordering the rows and columns of the graph matrix $\boldsymbol{G}$ according to the time to aborption vector $\boldsymbol{t}$ provides a way to visualize the amount and location of disagreement among rankings in P.

The visualization can be further enhanced by using grayscale tones to indicate not just the location but the *intensity* of the disagreement. The largest values in $\boldsymbol{G}$ appear darkest while the smallest values are shaded lightest.

TO DO. Aside with large, real examples of this way of visualizing P creating grayscale spy plots.

Computational Note. If n is large, computing $\boldsymbol{t}$ by $\boldsymbol{t} = (\boldsymbol{I} - \boldsymbol{T}_{11})^{-1}\boldsymbol{e}$ can be expensive due to the $O(n^3)$ matrix inversion. Fortunately, there are much more efficient ways to compute $\boldsymbol{t}$, particularly when $\boldsymbol{T}_{11}$ is sparse as it generally is for large n. The vector $\boldsymbol{t}$ can be computed with sparse matrix-vector multiplications.The iterative procedure below computes $\boldsymbol{t}$ using only sparse matrix-vector

multiplications and the storage of three $O(n)$ vectors. Because $\boldsymbol{T}_{11}$ is substochastic, the spectral radius is less than 1, which means that the Neumann series can be used in place of the inverse, i.e., $(\boldsymbol{I}-\boldsymbol{T}_{11})^{-1}=\boldsymbol{I}+\boldsymbol{T}_{11}+\boldsymbol{T}_{11}^2+\cdots$.

```
prevt=0;
temp=e;
t=e;
while norm(prevt-t,2)>.0001
    prevt=t;
    temp=T11*t;
    t=t+temp;
end
```

TO DO. Add Summary Box at end of this section with Algorithm for reordering $\boldsymbol{G}$ by $\boldsymbol{t}$ in order to visualize P

4.2 PATH PROBABILITIES FOR P

Each ranking in P is a path through the graph $\boldsymbol{G}$ and can be scored for its **path probability**. Applying this idea to our running example, produces the path probabilities listed above each ranking in P.

$$P=\left\{\overset{\frac{1}{216}}{\begin{pmatrix}3\\4\\1\\2\end{pmatrix}},\overset{\frac{1}{48}}{\begin{pmatrix}3\\4\\2\\1\end{pmatrix}},\overset{\frac{1}{216}}{\begin{pmatrix}2\\3\\1\\4\end{pmatrix}},\overset{\frac{1}{72}}{\begin{pmatrix}2\\3\\4\\1\end{pmatrix}},\overset{\frac{1}{72}}{\begin{pmatrix}3\\1\\4\\2\end{pmatrix}},\overset{\frac{1}{96}}{\begin{pmatrix}4\\2\\3\\1\end{pmatrix}}\right\}.$$

The ranking of [2 3 4 1] is the path from 0 to 2 to 3 to 4 to 1 to 5, which has path probability $(\frac{1}{3})(\frac{1}{2})(\frac{1}{2})(\frac{1}{3})(\frac{1}{2})=\frac{1}{72}$, whereas the ranking of [3 4 2 1] has a path probability of $\frac{1}{48}$. The path probability for the ranking of [2 3 4 1] is calculated by multiplying the probability of each link in the path, which are highlighted in bold red font in the $\boldsymbol{G}$ matrix below.

$$\boldsymbol{G}=\begin{array}{c}\\0\\1\\2\\3\\4\\5\end{array}\begin{array}{c}\begin{array}{cccccc}0&1&2&3&4&5\end{array}\\\begin{pmatrix}0&0&\mathbf{1/3}&1/2&1/6&0\\0&0&1/6&0&1/3&\mathbf{1/2}\\0&1/6&0&\mathbf{1/2}&0&1/3\\0&1/2&0&0&\mathbf{1/2}&0\\0&\mathbf{1/3}&1/2&0&0&1/6\\0&0&0&0&0&1\end{pmatrix}\end{array}.$$

For this example, of the rankings in P, the rankings of [3 4 1 2] and

[2 3 1 4] tie for the least probable and [3 4 2 1] is the most probable. Recall that this most probable ranking was the same ranking produced by the time to absorption vector $\boldsymbol{t}$.

4.3 FINDING OUTLIERS IN P

The previous sections presented ideas for finding rankings in P that either summarize or best represent the ranking information in P. In this section, we do the opposite. We are interested in the least representative ranking, i.e., the ranking that is the most different or most unusual. We call these unusual rankings **outlier rankings** because they are the farthest from the others.

Why might one want to find outliers in P? What value might they have? An outlier ranking is as equally optimal as any other ranking in P, yet unusual. One might think of them as minority or dissenting opinions.

Since we define an outlier ranking as farthest from the other rankings, we need to compute, for each ranking in P, its distance to every other ranking in P. There are several ways to measure the distance between two rankings, e.g., Spearman's footrule [] or Kendall's tau [] or NDCG [] or weighted variants of these []. Any is appropriate. Here we chose the weighted —- because it penalizes disagreements between two rankings that occur toward the top of the ranking.

EXAMPLE with weighted distance between two rankings below

Thus, we create a $p \times p$ matrix $\boldsymbol{O}$ of distance values between pairs of rankings. This matrix is symmetric, so only the $p(p-1)/2$ upper or lower triangular elements need be computed. Outlier rankings correspond to the rows in $\boldsymbol{O}$ with the smallest sum.