SUPPLEMENTARY MATERIALS: THE RANKABILITY OF DATA*

PAUL E. ANDERSON†, TIMOTHY P. CHARTIER‡, AND AMY N. LANGVILLE§

SM1. ACC example. *The relaxation with the LP's* $\mathbf{X}$ *matrix gives a* ***probabilistic*** *interpretation about which links to add. And the LP's* $\mathbf{Y}$ *matrix gives a* ***probabilistic*** *interpretation about which links to remove.* The following example demonstrates the previous statements with real data from the 2009 ACC college basketball season of games between $n = 12$ teams. The $\mathbf{D}$ matrix contains a 1 if team i beat team j in the majority of their matchups. We used the exact method to produce exact answers, in particular, $p = 165$ and the three matrices $\mathbf{P}_a$, $\mathbf{P}_d$, and $\mathbf{P}_>$. Compare these exact answers $\mathbf{P}_a$, $\mathbf{P}_d$, and $\mathbf{P}_>$ with the approximate answers provided by the LP matrices $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{D} + \mathbf{X} - \mathbf{Y}$, respectively.

$$\mathbf{X} = \begin{bmatrix} 0 & .54 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .53 & 0 & 0 & 0 & 0 & .22 & 0 & 0 & 0 \\ 0 & .23 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .26 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .30 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .42 & 0 & 0 & 0 & 0 & 0 & 0 & .09 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .27 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .25 \\ 0 & 0 & 0 & 0 & 0 & 0 & .91 & 0 & 0 & 0 & .75 & 0 \end{bmatrix} \quad \text{and}$$

$$\mathbf{P}_a = \begin{bmatrix} 0 & .60 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .60 & 0 & 0 & 0 & 0 & 0.20 & 0 & 0 & 0 \\ 0 & .20 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .20 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .24 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .52 & 0 & 0 & 0 & 0 & 0 & 0 & .03 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .20 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .24 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .15 \\ 0 & 0 & 0 & 0 & 0 & 0 & .96 & 0 & 0 & 0 & .84 & 0 \end{bmatrix}$$

Notice that the location of nonzeroes in $\mathbf{X}$ and $\mathbf{P}_a$ match exactly. Further, the values in these locations in $\mathbf{X}$ well-approximate the values in the exact $\mathbf{P}_a$. In fact, the deviation between the approximation and the exact value is at most .1. The same holds for the LP's $\mathbf{Y}$ matrix and $\mathbf{P}_d$.

*Submitted to the editors April 2018.

Funding: This work was funded by the —.

†Department of Computer Science, College of Charleston, SC 29401, USA (andersonpe2@cofc.edu).

‡Department of Mathematics and Computer Science, Davidson College, Davidson, NC (tichartier@davidson.edu).

§ Department of Mathematics, College of Charleston, SC 29401, USA (langvillea@cofc.edu).

$$\mathbf{Y} = \begin{bmatrix} 0 & 0 & .22 & 0 & 0 & 0 & 0 & 0 & .25 & 0 & 0 & .27 \\ .54 & 0 & .22 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .38 \\ 0 & 0 & 0 & .26 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .48 \\ 0 & .53 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .49 \\ 0 & 0 & 0 & 0 & 0 & 0 & .42 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & .64 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & .67 & 0 & .27 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & .33 & 0 & 0 & 0 & 0 & 0 \\ 0 & .22 & .52 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .28 \\ 0 & 1 & 0 & 0 & .3 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .73 & .62 & .52 & .51 & 0 & .36 & 0 & 0 & .72 & 0 & 0 & 0 \end{bmatrix} \quad \text{and}$$

$$\mathbf{P}_d = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .20 & 0 & 0 & .19 \\ .60 & 0 & .20 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .36 \\ 0 & 0 & 0 & .20 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .48 \\ 0 & .60 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .53 \\ 0 & 0 & 0 & 0 & 0 & 0 & .52 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & .73 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & .73 & 0 & .24 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & .27 & 0 & 0 & 0 & 0 & 0 \\ 0 & .20 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .24 \\ 0 & 1 & 0 & 0 & .24 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .81 & .64 & .52 & .47 & 0 & .27 & 0 & 0 & .76 & 0 & 0 & 0 \end{bmatrix}$$

Once again, the location of nonzeroes in $\mathbf{Y}$ and $\mathbf{P}_d$ match exactly. Further, the values in these locations in $\mathbf{Y}$ well-approximate the values in the exact $\mathbf{P}_d$. In fact, the deviation between the approximation and the exact value is at most .1. The same holds for the comparison between the LP's $\mathbf{D}+\mathbf{X}-\mathbf{Y}$ and $\mathbf{P}_>$. One more observation comes from this $n = 12$ ACC example. The number of nonzeroes in $\mathbf{X}$ (and $\mathbf{P}_a$) is 14 while the number of nonzeroes in $\mathbf{Y}$ (and $\mathbf{P}_d$) is 36. This means that almost three times as many link deletions as link additions are required to transform this data into a perfect dominance matrix. In other words, the data consists of more *inconsistent ranking information* (indicated by link deletions) than *missing ranking information* (indicated by link additions).

SM2. LP accuracy experiments. This section contains further experiments measuring the accuracy of the LP approximate method. We assess the accuracy of the LP's approximations in two ways: (1) with small n for general graphs (see Table SM1) and (2) with larger n for graphs with special structure (see Fig. SM1). As described earlier, we know the exact answers for $\mathbf{P}_a$, $\mathbf{P}_d$, $\mathbf{P}_>$ for completely connected, empty, and cyclic graphs of any size n.

Table SM1 displays the average accuracy of the LP approximations $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{D}+\mathbf{X}-\mathbf{Y}$ compared to the exact results in the $\mathbf{P}_a$, $\mathbf{P}_d$, and $\mathbf{P}_>$ for thousands of randomly generated small graphs. Artificial random graphs with varying degrees of sparsity are created by randomly placing a fixed number of links l in the $n^2 - n$ link locations of an empty graph. The bigger l is, the denser the random graph. We used the exact method to compute the full set P, and thus, the exact matrices $\mathbf{P}_a$, $\mathbf{P}_d$, and $\mathbf{P}_>$. Notice that the locations of nonzeroes in the LP results match the location of nonzeroes in the exact $\mathbf{P}$ matrices for nearly all graphs. The few instances in which the match is nearly perfect is due to numerical precision of the interior point solver of the LP algorithm. Interior point solvers converge within some tolerance of the optimal solution and thus they do not arrive exactly at the centroid of the optimal solution. Not only do the locations of nonzeroes match, but the values are very close as shown by the `relative deviation`, `avg absolute dev`, and `avg max abs dev` sections of

TABLE SM1
Accuracy of LP results compared to exact results

Approximate vs. Exact	Size	X vs. P_a	Y vs. P_d	$D + X - Y$ vs. $P_>$
nz location match	5	100.00%	100.00%	100.00%
	6	100.00%	100.00%	100.00%
	7	100.00%	100.00%	100.00%
	8	100.00%	100.00%	100.00%
	9	100.00%	100.00%	100.00%
	10	100.00%	99.38%	99.38%
relative deviation	5	0.03	0.03	0.02
	6	0.03	0.04	0.02
	7	0.06	0.06	0.04
	8	0.06	0.06	0.04
	9	0.08	0.08	0.05
	10	0.08	0.09	0.05
avg absolute dev	5	0.00	0.00	0.01
	6	0.01	0.01	0.01
	7	0.01	0.01	0.02
	8	0.01	0.01	0.02
	9	0.01	0.01	0.02
	10	0.01	0.01	0.03
avg max abs dev	5	0.02	0.02	0.03
	6	0.04	0.04	0.05
	7	0.07	0.07	0.09
	8	0.08	0.09	0.10
	9	0.11	0.11	0.13
	10	0.13	0.13	0.15

Table SM1. For example, in the worst case, for a $n = 7$ graph, the maximum difference in any element in the approximate $\mathbf{X}$ versus the exact $\mathbf{P}_a$ is .07.

The next table (Figure SM1) extends these findings to graphs with bigger n by examining graphs with special structure, those for which we know the exact answers for $\mathbf{P}_a$, $\mathbf{P}_d$, $\mathbf{P}_>$. The table in Figure SM1 shows the accuracy of the LP's approximation to $\mathbf{P}_a$, $\mathbf{P}_d$, $\mathbf{P}_>$ for increasing values of n. For example, the LP's output matrix $\mathbf{X}$ approximates, without the need to generate the full set P, the matrix $\mathbf{P}_a$ built from P. The first column of Table SM1 reports the percentage of elements in the LP's $\mathbf{X}$ matrix that match the locations of nonzero elements in the exact $\mathbf{P}_a$. Recall that $\mathbf{P}_a(i, j)$ is the percentage of members in P that recommend adding a link from i to j. Thus, with 100% match, the LP's $\mathbf{X}$ is identifying these very same links.

The LP's $\mathbf{X}$ *matrix is not just identifying these links, it is also approximating the weight or percentages very well as measured by the deviation between the approximate* $\mathbf{X}$ *and the exact* $\mathbf{P}_a$*.* For example, the `relative deviation` between $\mathbf{X}$ and $\mathbf{P}_a$, $\|\mathbf{X} - \mathbf{P}_a\|/\|\mathbf{P}_a\| = .03$ for $n = 25$, showing how numerically close the two matrices $\mathbf{X}$ and $\mathbf{P}_a$ are. The columns labeled `absolute deviation` show element-wise deviations between the approximate matrix $\mathbf{X}$ and the exact matrix $\mathbf{P}_a$. In this case, for the $n = 25$ cyclic matrix, the average deviation of an element in $\mathbf{X}$ from the exact value in $\mathbf{P}_a$ is .01, while the maximum value of this absolute deviation is .04, showing, again, that the two matrices $\mathbf{X}$ and $\mathbf{P}_a$ are very close.

These accuracy experiments were run in MATLAB on a Macintosh laptop with a 2.8GHz processor and 16GB of memory. Notice that LPs with $n = 75$ take just a few seconds to produce results for these three types of structured graphs. These results are encouraging–in seconds, we can get very good approximate rankability information, making graphs with n in the hundreds achievable in real-time.

FIG. SM1. ***Table comparing the LP's matrices to the exact matrices built from the set P.*** *The LP matrices give excellent approximations to both the location of nonzero elements and the values of these elements for cyclic, completely connected, and empty graphs of increasing size. Recall that the LP method does not provide a value for p and, thus, r either.*

Cyclic	***n=25***				***n=50***				***n=75***			
Approximate vs. Exact	***nz location match***	***relative deviation***	***abs. dev. avg***	***max***	***nz location match***	***relative deviation***	***abs. dev. avg***	***max***	***nz location match***	***relative deviation***	***abs. dev. avg***	***max***
X vs. P_a	100%	.03	.01	.04	100%	.01	.007	.02	99.9%	.008	.005	.01
Y vs. P_d	96%	.96	.003	.04	99%	.97	10^{-4}	.02	99.9%	.98	10^{-4}	.01
D+X-Y vs. $P_>$	99%	.02	.01	.04	99%	.01	.007	.02	99.2%	.008	.005	.01
$\tilde{p}$ vs. p	$\tilde{p} = 600$	$p = 10^{25}$			$\tilde{p} = 2450$	$p = 10^{64}$			$\tilde{p} = 5550$	$p = 10^{109}$		
LP time	.14 sec				1.56 sec				10.31 sec			

Completely Connected/Empty	***n=25***				***n=50***				***n=75***			
Approximate vs. Exact	***nz location match***	***relative deviation***	***abs. dev. avg***	***max***	***nz location match***	***relative deviation***	***abs. dev. avg***	***max***	***nz location match***	***relative deviation***	***abs. dev. avg***	***max***
X vs. P_a	100%	10^{-11}	10^{-11}	10^{-11}	100%	10^{-14}	10^{-14}	10^{-14}	100%	10^{-14}	10^{-14}	10^{-14}
Y vs. P_d	100%	10^{-11}	10^{-11}	10^{-11}	100%	10^{-14}	10^{-14}	10^{-14}	100%	10^{-14}	10^{-14}	10^{-14}
D+X-Y vs. $P_>$	100%	10^{-11}	10^{-11}	10^{-11}	100%	10^{-14}	10^{-14}	10^{-14}	100%	10^{-14}	10^{-14}	10^{-14}
$\tilde{p}$ vs. p	$\tilde{p} = 600$	$p = 10^{25}$			$\tilde{p} = 2450$	$p = 10^{64}$			$\tilde{p} = 5550$	$p = 10^{109}$		
LP time	.02 sec				.76 sec				6.28 sec			