Math 160 - Project 1

Sangwon Yoon, Ryan Tjoa and Ahmed Mahmoud

April, 18th 2018

$$max\begin{cases} H_{i-1,j-1}+1 \\ H_{i-1,j}-2 \\ H_{i,j-1}-2 \\ \end{cases} \tag{1}$$

Problem No.2

Using the plurality vote method, we rank each candidate by the amount of first place votes they receive. The winner of the election is Bernie Sanders with 96 first place votes.

Using the average rank method, we rank each candidate by assigning points from 1 to n, where 1 is the voter's least favorite candidate and n the voter's most favorite candidate, where n is the total number of candidates. After collecting all the data from the all the voters, we take the sum of all the points for each candidate and divide by the total number of voters to get the average rank of each candidate. By this method, Hilary Clinton wins the election with a rank of 3.8583, slightly above Bernie Sanders who has a rank of 3.7417.

Using the Borda count method, we rank each candidate by assigning points from 0 to $n-1$, where 0 is the voter's least favorite candidate and $n-1$ is the voter's most favorite candidate, where n is the total number of candidates. After collecting all the data from the voters, we take the sum of all the points for each candidate, and the candidate with the highest sum (the Borda score) is the winner. By this method, Hilary Clinton is the winner with a Borda score of 686, and Bernie Sanders is in second with a Borda score of 658.

Using the W-Borda count method, we rank each candidate just like how we do in the Borda count method. However, the assigned points are given by a vector W, where $w_1 \geq w_2 \geq ... \geq w_n$, with w_1 being the points assigned to the candidate in first place and w_n to the candidate in last place. Using the W vector values [1 1/2 1/3 1/4 1/5] we can see that Bernie Sanders wins the election with a score of 148.5 and Hilary Clinton with a score of 147.9. However, if we change the W vector values to 5 4 3 2 1], Hilary Clinton wins the election with a score of 926 and Bernie Sanders having a score of 898.

Using the Pagerank algorithm, we rank each candidate by calculating the A matrix. We do this by creating a $5x5$ matrix for each voter denoting their possible choices, and mark each candidate with a value of 1 if the candidate on the given row loses to the corresponding candidate on the given column.

By taking the summation of all the voters' matrices, and dividing each column by the respective column sum, we get the A matrix. We then calculate $\widetilde{A}$ using the equation given in class, and from there we compute $Ar = r$ using the power method. By this method, we get that Hilary Clinton wins the election with a score of 0.24981 while Bernie Sanders has a score of 0.246.

We see that depending on the vote method chosen, we either see Hilary Clinton or Bernie Sanders as the winner. By plurality method, we see that Bernie Sanders is the winner because he has the most first place votes. However, by the average rank method, Borda count method, and Page Rank method, we see that Hilary Clinton is the winner. This tells us that to a portion of the voters, Bernie Sanders was the top choice, and to the other portion, Bernie Sanders was more towards the less favorable side, while Hilary Clinton had most of her votes towards the favorable side.

Using the W Borda method, we see that manipulating the W vector can give us two different winners, either Hilary Clinton or Bernie Sanders. A vector with a larger point margin between the first and last candidates usually results in Hilary Clinton winning, while a vector with a smaller point margin between the first and last candidates usually results in Bernie Sanders winning.

In our opinion, Page Rank is the fairest method to count votes because the margin in between each candidate is the smallest. This will give us the best representation of how close the election is, and if there is a need for a recount.

Plurality	Average	Borda	W-Borda1	W-Borda2	Page Rank
BS (96)	HC (3.8583)	HC (686)	HC (926)	BS (148.05)	HC (0.24981)
HC (85)	BS (3.7417)	BS (658)	BS (898)	HC (147.9)	BS (0.246)
DT (44)	DT (2.6958)	DT (407)	DT (647)	DT (101.15)	DT (0.19265)
TC (8)	JK (2.5375)	JK (369)	JK (609)	JK (78.1167)	JK (0.16525)
JK (7)	TC (2.1667)	TC (280)	TC (520)	TC (72.7833)	TC (0.14629)

Figure 1: Ranking of the candidates using different methods along with the each method rating between parenthesis. W-Borda1 corresponds to the weight vector of [5 4 3 2 1] while $W - Borda2$ is for weight vector $[1\, \frac{1}{2}\, \frac{1}{3}\, \frac{1}{4}\, \frac{1}{5}]$.

Problem No.3

(i) Using SVD, we can approximate the matrix G by a rank-one matrix. A rank-one matrix consists of a set of linearly *dependent* columns. Using the column associated with largest singular value i.e., u_1 from the left singular vectors and v_1 from the right singular vectors and approximate G as $u_1 v_1^T$, the result is an approximation that captures the important feature of what G represents i.e., the questions difficulty (or easiness since large number means students can answer this question with ease). Since the approximation matrix is rank-one, that means by picking any column, we get a rough idea about the difficulty of the question such that the entry in the column associated with higher value represents the easiest question and second largest represents second easiest question and so on.

(ii) Using the score differentials gives a good indications what is the easiest and most difficult questions are since this margin or difference shows how one student can perform on one question compared to the other. Using least square to solve the system of equation given by Massey's, we can reach the best rating for the system. We note here that the model given (how to construct the matrix) has a great impact on the final answer Massey's method gives. It also possible to construct the matrix differently and the answer might be different.

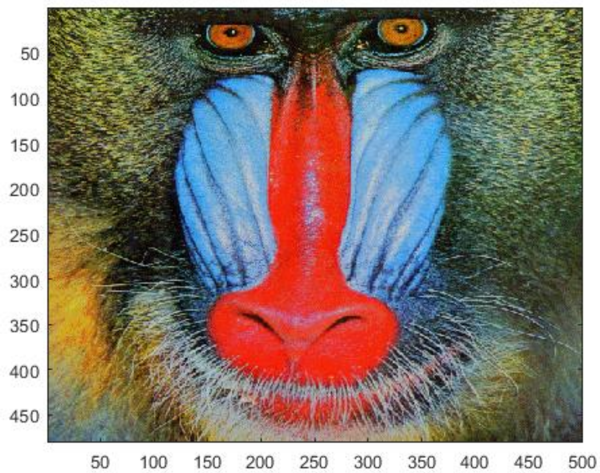
(iii) Shown below the score or rating using the three methods. We can see that all methods are consistent about what is the hardest question which give us results with confidence. Where Massey's and Colley's agree in their ratings, SVD method deviates for subsequent questions. This could be due to the fact that SVD does not use all the information of the matrix as it approximate it and give rough estimate based on this approximation.

SVD	Massey's	Colley's
Q.7 (-0.1191)	Q.7 (-0.81567)	Q.7 (0.32192)
Q.4 (-0.080687)	Q.2 (-0.71889)	Q.2 (0.44977)
Q.2 (-0.056591)	Q.3 (-0.10599)	Q.3 (0.4726)
Q.6 (0.052502)	Q.4 (0.087558)	Q.4 (0.50913)
Q.5 (0.054354)	Q.5 (0.34562)	Q.5 (0.54338)
Q.3 (0.084453)	Q.6 (0.57143)	Q.6 (0.57306)
Q.1 (0.092386)	Q.1 (0.63594)	Q.1 (0.63014)

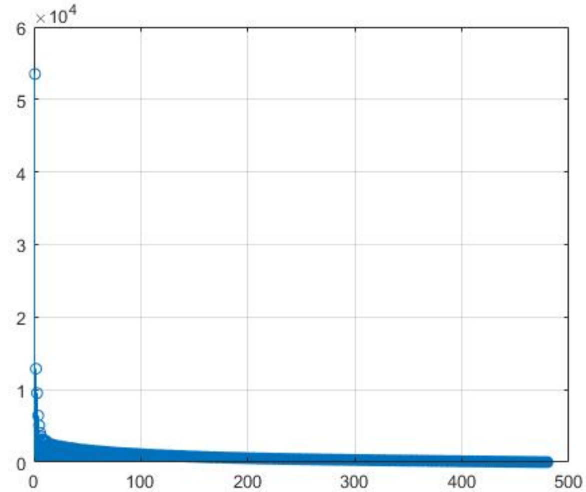
Figure 2: Ranking of the questions using SVD, Massey's network and Colley's method starting from the most difficult to easiest. The score or rating given by each method is shown between parenthesis for each question.

Problem No.4

Figure 3 shows the image along with the distribution of its singular value. The distribution shows a single spike which rapidly smooth out. We think this is due to the greater intensity along the nose of the *mandrill*.



(a) *mandrill* image



(b) distribution of singular values

Figure 3: The image (left) along with the distribution of its singular values (right).

Figure 4 shows approximation of the image using rank k approximation based on SVD. As the rank increases, more details appears as we are extracting more details using the largest singular value then the second largest and so on.

Figure 5 shows the relative and absolute error of the rank k approximation. As it was suggest from Figure 4, increasing the rank reduces the error. While increasing the rank from 1 to 6 decreased the error by third, moving to higher ranks does not scale the same which suggests (at least for this example) that using fewer columns of SVD can approximate the image well.

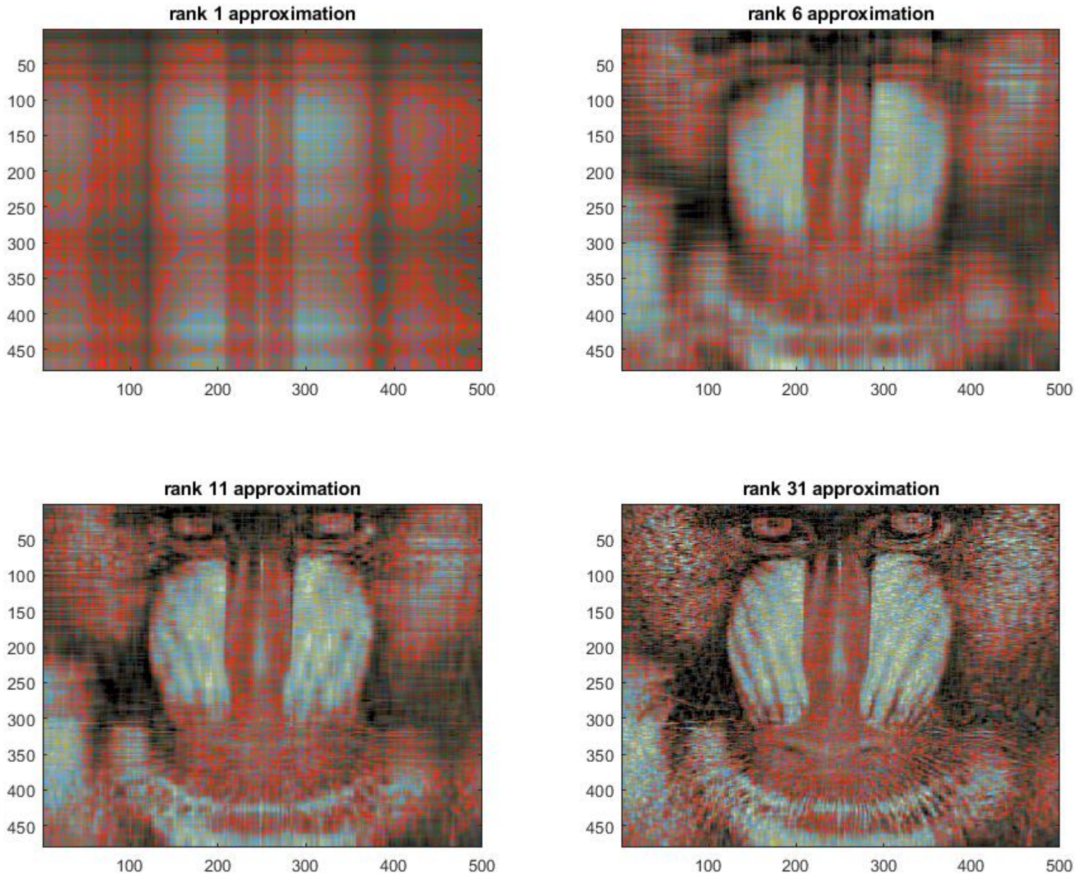


Figure 4: Four different approximation for the image based on the rank.

	Absolute Error	Relative Error
Rank 1 approximation	12891.1649	$1.2699e^{-15}$
Rank 6 approximation	3537.8836	$6.4268e^{-16}$
Rank 11 approximation	2820.0648	$6.4502e^{-16}$
Rank 31 approximation	1985.461	$4.5808e^{-16}$

Figure 5: The absolute error (norm) along with relative error of the approximation of image based of the rank.