

The student ID of the student whose paper you are grading *

3032130362



Readability and grammar of written report (5 points) *

	1	2	3	4	5	
Difficult to read and/or poor grammar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Clearly written and excellent grammar

Level of written detail on comparison of R and C++ implementation and runtime (3 points) *

	0	1	2	3	
Did not write about a comparison of the R and C++ implementation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Wrote a detailed comparison between the R and C++ implementations



Review the code written by the author. If you aren't sure of the correctness of the implementation, that's fine, just give a grade and say so in the comments.

Correctly coded the parallelization of k-means and pairwise similarity in R/C++ (3 points) *

	0	1	2	3	
incorrect implementation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	seems correct to me

Comments on implementation of parallelization or the similarity measure?

Efficiency and practicality of R and C++ code (3 points) *

	1	2	3	
inefficient (e.g. repeated computations unnecessarily, saved objects unnecessarily, etc)	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	very efficient and practical

Suggestions for improving *efficiency* of R and/or C++ code *

could use the same C++ "integrator" code for R implementation of correlation calculation instead of forming the matrix

Does the author satisfy the following code readability requirements? (3 points) *

- ☐ Consistent spacing before and after variable assignment and addition symbols (" = ", " + "), and after commas (" , ")
- ☐ No line of code exceeds 80 characters
- ☒ Consistent variable naming (words always separated by one of "_" or ".")

Clarity of variable names (2 points) *

	0	1	2	
variable names are unclear and meaningless (eg `df`, `x`, `data2`, etc)	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	variable names are helpful and unambiguous

Quality of code comments (2 points) *

	0	1	2	
there are almost no comments	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	the comments explain clearly what is being done and why

Suggestions for improving *readability* of R code *

some commas don't have spaces after them, also the "stabilize" function has an unintuitive name

Did the student provide all code necessary for recompiling their results AND report (note: you do not have to actually reproduce their report) (2 points) *

	0	1	2	
Incomplete code or no .Rnw/.Rmd file provided	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Everything was provided

Clarity of folder structure (2 points) *

	0	1	2	
The folder structure was very confusing	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	It was clear what each file corresponded to and there were no surplus files floating around

Optional comments on folder structure and files provided (please provide comments if you docked points for any reason)

Figures

Correctly produced Ben-Hur-type figures (3 points) *

	0	1	2	3	
Did not provide a figure like Ben-Hur	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Figures look correct

If the Ben-Hur figures do not look correct, what is wrong?

Quality of Ben-Hur Figure 3 replication figures (3 points) *

	0	1	2	3	
Did not provide a figure like Ben-Hur	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Provided clear and visually appealing figures

Discuss one (or more) things that you liked about the author's Ben-Hur figures *

the colors on the CDF plot are nice

Discuss one (or more) things that could be improved for the author's Ben-Hur figures *

use the same x axis for each of the histograms, and also larger bin size, they are quite spiky

Justification of conclusions drawn from the Ben-Hur-type figures (3 points)

*

	0	1	2	3	
Did not write about any conclusions drawn from the figures	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Clearly outlined interpretations of the figures and drew reasonable conclusions (e.g. found $k = 3$, or some other value, is the best and provides reasons why)

Comments on the conclusions and interpretations of the Ben-Hur type figures *

it does look like three clusters gives the "most concentrated" Jaccard histogram, but it takes a some time to make sense of the different x axes and the Jaccard coefficient does not compare well between different choices of k , for example, $k = 2$ has much higher values than $k = 3$

Conclusion



Provide concluding comments

One or more things that you thought was well done overall *

i really like the table of runtime comparisons

One or more things that could be improved upon overall *

talk more about the data and if the conclusion for the "best" k seems intuitive or justified from the data

Any other comments that you would like to add?

This form was created inside of UC Berkeley.

Google Forms