COVID-19 Risk Assessment: Preparation and

Response

Mayuresh Hooli   
Technology & Solutions  
Ericsson Inc.Plano, USA  
mayur.hooli@ericsson.com

Mason Jerome  
ECEN  
Texas A&M UniversityCollege Station, USA  
mmjerome15@gmail.com

*Abstract*— COVID-19 has been declared a global pandemic. Though the mortality rate is slightly lower than 5%, the rate of spread is unprecedented since the Spanish Flu in the early 20th century. The registered number of cases in USA was 174,684 as of April 1, but this number is set to see a sharp increase as the rate of cases has jumped significantly. With a limited number of test kits available, we need to identify the counties that are hotspots for the spread of the virus for proper allocation. Once the hotspots are identified, we need to identify the rate of spread to neighboring counties via our Machine Learning and Recommendation System techniques. If these critical parameters can be identified, then there will be some measures in place that will be crucial in fighting this virus. This approach helps to optimize the use of the kits and addresses the most important issues we have right now through a disaster management approach. As time passes, we can evaluate the results of our approach with the actual value of cases in each county in Texas. We extract some meaningful features out of public census datasets and try to identify the hotspots. We will use data analytics and statistical analysis to study the effect of each indicator on spreading the virus and predicting its spread into hotspots. This idea can help local and state governments to logistically optimize their resource allocation.

Keywords—COVID-19, Ranking (Statistics), Clustering, Information Retrieval.

# Introduction

Coronavirus disease 2019 (COVID-19) or Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a disease closely related to SARS-CoV which was responsible for the SARS epidemic of 2003 [1]. Coronavirus belongs to a group of viruses that infects various animals like dogs and bats. The name coronavirus originates from the bulb-tipped spikes that project from the surface of the virus and give the appearance of a corona surrounding it. The coronavirus infection tends to play out in two ways: either as an injection in the lungs that would cause common cold or an infection in the guts that would cause diarrhea. This starts out in the lungs like a common cold, but then wreaks havoc on the immune system in some cases resulting in long-term lung damage amounting to death. While this bears similarities to the SARS-CoV, there are over 6000 mutations sprinkled around [2].

It is difficult to determine the origin of COVID-19, though it is improbable that it emerged from laboratory manipulation of SARS-CoV. Two scenarios can plausibly explain the origin of COVID-19: (i) natural selection in an animal host before zoonotic transfer; and (ii) natural selection in humans following zoonotic transfer.

## Natural selection in an animal host before zoomotic transfer

The early cases of COVID-19 were related to the Huanan market in Wuhan which postulates an animal source present at this location. As the SARS-CoV-like virus in bats was similar to the SARS-CoV-2 virus, it is likely that bats were the initial hosts of this virus. Although this is ~96% similar to SARS-CoV-2, recent studies have suggested that its spike diverges in RBD, which suggests it may not bind efficiently to human ACE2. Another potential source is the Malayan pangolin illegally imported in to the Guangdong province which contain coronaviruses similar to SARS-CoV-2 [3].

## Natural selection in humans following zoonotic transfer

This is paragraph 2.

The COVID-19 Management is an intelligent system which identifies the critical areas (case study: Texas) where certain actions need to be taken. This system also ranks the counties based on the risk they pose to other counties for spreading the virus. This novel idea is based on implementing clustering alongside the PageRank algorithm. Clustering identifies 4 clusters based on the data set we prepared. PageRank algorithm uses the clustering prediction to identify which counties risk spreading the virus further. There is a shortage of medical personnel and testing kits, the categorization of counties helps in effective allocation of resources. The prediction of future affected counties will help in better preparedness to mitigate the virus. Ericsson is a pioneer in the field of communication. The communication technology developed by Ericsson would help understand the travel history of a county on an average to better prepare for the coming weeks.

# Proposed Method

The idea is more useful for state and local government in terms of optimizing resources in this difficult time. The Disaster Management tool which can get funding from the government. The government could benefit from its analysis to better control the distribution of test kits and implement policies to reduce the spread of the virus. People are interested to get a solution to see how safe different counties are. Ericson can show that its AI/ML know-how is valuable at these difficult moments.

## Novelty

Policymakers around the world is trying to figure out how to fight Coronavirus. The novel approach of using Clustering and Page Rank together can help them understand where the risk lies. Following the results given by this algorithm would help manage the fight against the pandemic.

*B. Business Values*

Ericsson’s infrastructure would help predict the spread better by using the cell phone data to understand the travel history. This idea is useful for the Government which would help Ericsson get more Government contracts.

*C. Page Rank Algorithm*

...

*D. Clustering*

Clustering is a technique that divides data points into different groups based upon their features. Using this method provides clear groupings for the counties that lets us classify them into as many clustered segments as chosen. For the purpose of this dataset, we chose to delineate the counties into 4 different clusters with K-means clustering. This algorithm finds datapoint closest to each assigned cluster center and iterates, updating the cluster center locations until error is minimized to find the best cluster assignments. Upon arriving at the final clusters, the Page Rank algorithm is used to determine the severity of COVID-19’s impact upon each cluster of counties.

*E. Data*

We extract some meaningful features out of public census datasets and try to identify the hotspots. We will use data analytics and statistical analysis to study the effect of each indicator on spreading of the virus and predicting the spread into neighboring hotspots. The census data that we use for hotspot identification has not been used before.

# Prepare Your Paper Before Styling

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

## Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

## Units

* Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
* Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
* Do not mix complete spellings and abbreviations of units: “Wb/m2” or “webers per square meter”, not “webers/m2”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.

Identify applicable funding agency here. If none, delete this text box.

* Use a zero before decimal points: “0.25”, not “.25”. Use “cm3”, not “cc”. (*bullet list*)

## Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

*a**b* 

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

## Some Common Mistakes

* The word “data” is plural, not singular.
* The subscript for the permeability of vacuum **0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
* In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
* A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
* Do not use the word “essentially” to mean “approximately” or “effectively”.
* In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
* Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
* Do not confuse “imply” and “infer”.
* The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
* There is no period after the “et” in the Latin abbreviation “et al.”.
* The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [7].

# Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Authors and Affiliations

**The template is designed for, but not limited to, six authors.** A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors.

## Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

## Figures and Tables

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord “Format” pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.

I. INTRODUCTION

*B. Contribution*

The COVID-19 Management is an intelligent system which identifies the critical areas (case study: Texas) where certain actions need to be taken. This system also ranks the counties based on the risk they pose to surrounding counties for spreading the virus. This novel idea is implements clustering alongside the PageRank algorithm. Clustering identifies 4 key clusters based on the dataset we prepared. PageRank algorithm uses the clustering prediction to identify which counties risk spreading the virus further. The prediction of future affected counties will help in better preparedness to mitigate the virus.

Identify applicable funding agency here. If none, delete this.

The communication technology developed by Ericsson would help understand the travel history of a county on an average to better prepare for the coming weeks

*C. Related Works*

...

II. PROPOSED METHOD

*A.*

Demo

County Name

Population Households

COVID-19 Cases

Population Distribution

COVID-19 Deaths

Population Density

Stay At Home Order

Clustering

Stay Inside **DANGER!**

Health Index

Travel Index

Enjoy Spring a bit

Don’t Enjoy Spring

Page Rank Algorithm Risk of Spreading the virus to other Counties

Here, we predict based on the counties. This can be scaled to a larger level for states and countries as well. But with a more specified set of census data, we can predict better based on the counties. Multiple data sets are combined to get the final, aggregated data set required for this Disaster Management tool such as census, cleanliness, and COVID-19 data.

The census data consists of the Geo ID of the county, FIPS, county name, the number of households, and the total male and female population as well as the population based on the percentages, males per 100 females and the same data distributed by age groups for every 5 years. We combine the total population categories for the age ranges of 0-15, 15-65 and 65+ for the data set. We do not consider the gender ratio for the data set.

The next data set we include is the health index. The health index is based on a data set obtained from the healthiest communities’ data set on US News. This data set contains information about the population health, equity, education, economy, housing, food nutrition, environment, public safety, community vitality, and infrastructure of the county. We aggregate all this data to calculate the cleanliness index of the counties to use in the data set. As we do not need the rest of the data now, we delete those columns and only use the aggregated column.

Next, we obtain the population density from the land area. This is used due to decreased social distancing and higher rates of contact between people in places where the population density is higher. The population density is aggregated for the county rather than considering the concentration in a few places that is seen in various counties.

The next parameter considered is the travel index. The travel data is restricted; hence, we try different methods to derive travel history. We assume that every person with a passport has traveled. From this assumption, we observe that there is a correlation between the travel data and the per capita income. There is also a correlation between the travel data and

education. Using these parameters, we calculate the travel index.

The next parameter we consider is the COVID-19 cases and deaths in each county. We also use the COVID-19 cases and deaths by population and household. Since these numbers are very small, we normalize them to get something we can use. We also use the COVID-19 stay at home labels. Since many of the Texas counties are rural counties with a small population density, people tend not to interact much with each other. We give the value 0 for counties that haven’t implemented stay-at-home orders and 1 for counties that have implemented the stay-at-home orders.

*F. Experimental Results*

Using the parameters detailed above, we implement the k-means clustering algorithm. We group the counties into 4 different clusters. The following diagrams give the results of the clusters in table I.

*G. Spread Identification*

Now we have identified the counties which are hotspots for the COVID-19. However, there may be more counties that are at risk of becoming COVID-19 hotspots in the future. We base this on the proximity of the counties to each other. The counties that have a high concentration of COVID-19 cases may have an adverse impact on other counties and raise their number of COVID-19 cases. To identify the threat posed by other counties to the county in question, we use the PageRank algorithm. To use the PageRank algorithm, we use the matrix of neighbors. But as we see above, not all the neighbors are equal. Thus, we replace the value 1 with an addition of cases per unit area and deaths per unit area and then take a square root of them. Then we normalize this value to get the values in the matrix. We apply a PageRank algorithm to this matrix. Based on this matrix, we observe that the Dallas county has the highest risk of being infected even further while Wheeler County is at the lowest risk.

Fig. 1. The general view of our proposed methodology.

Texas Counties Aransas Archer Austin Bandera Baylor Elpaso Hidalgo Wichita Nueces Cameron Travis Denton Collin Fort Bend Galveston Bexar Dallas Harris Tarrant

..

COVID-19

cases reported 2

0 2 0 0 40 28 28 22 20 200 165 134 119 70 157 488 526 139

Labels  
 Enjoy Spring a bit Enjoy Spring a bit Enjoy Spring a bit Enjoy Spring a bit Don’t Enjoy Spring Don’t Enjoy Spring Don’t Enjoy Spring Don’t Enjoy Spring Don’t Enjoy Spring Stay Inside Stay Inside Stay Inside Stay Inside Stay Inside Stay Inside Danger Danger Danger Danger

TABLE I EXPERIMENTAL RESULTS

III. FUTURE WORK

We plan to collect the cell phone data to better understand the travel history and aggregate based on county as a whole Extend this algorithm to the whole country by collecting data from all counties. We will use the Geographic Information System to identify the zones on the map for better visualization. We will combine model with time series analytic. We will create a similar model for pandemic response and disaster response to help manage similar situations in the future.

...

IV. CONCLUSION

REFERENCES

* [1] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7438984/>
* [2]