Project 1: What can the Logistic Map explain about COVID dynamics?

Colin Milhaupt
Graduate School of Computer Science
University of New Mexico
Albuquerque, United States
cmilhaupt@unm.edu

Gregory Jacobus
Graduate School of Computer Science
University of New Mexico
Albuquerque, United States
gjacobus@unm.edu

***Abstract*—The logistic map and related information theoretic metrics can be leveraged as tools to uncover new methods that model COVID cases in real populations. We study the relationship of new COVID cases among a selection of states in the southwestern United States to answer the question of whether or not trends in one state are a predictor of new cases in another state. Using transfer entropy and mutual information, we find that number of cases in one state may be able to help predict cases in another, but some states share too similar of data for any useful predictions. We then look a look at the average transfer entropy between states by variant in an attempt to characterize the evolution of the virus over time. We find that the Delta and Omicron variants of COVID had significantly higher transfer entropy between southwestern states compared to their Alpha predecessor, and discuss some possibilities that could lead to this result.**

***Index Terms*—COVID-19, logistic map, entropy, transfer entropy, mutual information**

I. Introduction

The Logistic Map is a popular model that can be used to approximate population growth, the spread of diseases, and much more. Its equation can be seen in (1).

$$x_{t+1} = Rx_t(1 - x_t) \tag{1}$$

In this equation, x_t is the value (e.g. population value) at time t and R is the population growth rate (i.e. $birth\_rate - death\_rate$) [4]. The $1-x_t$ serves to push the population back down within the model, attempting to simulate a population approaching carrying capacity when x_t is large. Furthermore, the logistic map can be used to determine when a model has sensitive dependence on initial conditions. This would mean that a small change to the initial population would cause drastic changes in future population after some number of time steps. As we will see, the logistic map exhibits sensitive dependence on initial conditions when it is in the chaotic regime, typically when $R > 3.5$.

Many measures of causation have been proposed over the last several decades. The one we'll be using to examine any potential causation between time-series is transfer entropy, proposed by Schrieber in 2007 and defined as:

$$T^{(k)}_{Y->X} = \sum_n p(x_{n+1}, x^{(k)}_n, y^{(k)}_n) \log[\frac{p(x_{n+1}|x^{(k)}_n, y^{(k)}_n)}{p(x_{n+1}|x^{(k)}_n)}] \tag{2}$$

where $p(x_{n+1}|x_n, ..., x_{n-k-1})$ describes a transition probability where each state x_{n+1} of the process is dependent on the last k-states but is independent of the state x_{n-k} and all previous states [6]. In its essence, transfer entropy describes the deviation from the expected entropy of two completely independent processes [7].

The chaotic regime of the logistic map historically implies a bottom-up causation. The line of thinking goes: if we can understand and model the position and velocity of every particle in the universe at time t, then we can predict the state of the universe at time $t+1$. Though, from studying fields like quantum mechanics, we know this isn't true and nature is in fact inherently probabilistic. When modeled with the logistic map, Walker et al. argue that there is a switch in causation from bottom-up to top-down for values of epsilon between 0.2 and 0.7, where epsilon is the global coupling strength to the instantaneous dynamics of a mean-field, m_n [7].

We test this claim by studying the values of epsilon between 0.2 and 0.3, where Deterding and Wright claim the mean-field seems to have an influence on the dynamics of the system [1]. Though they acknowledge the high uncertainty in this area, we will take a closer examination.

We conclude by applying the logistic map and measures of transfer entropy to study the dynamics of population growth of COVID-19 infection in the southwestern United States. We start with a localized view comparing 2 states with similar populations, socioeconomic backgrounds, and political leaning. We then expand our view to look at southwestern states that border each other and have a higher chance of experiencing inter-state migrations between populations.

II. Methods & Results

A. Logistic Map, Entropy, and Mutual Information

1) Logistic Map: To understand more about the logistic map, we plot it in a couple different ways, as shown in Fig. 1. For plotting, we use `Python3` and its `matplotlib` library, using code based on [8]. First, we plot the logistic map in

the non-chaotic regime, using $r = 2.8$, as shown in Fig. 1a. We set the initial population (y_0) to two different values, 0.1 and 0.11, to help determine how sensitive the logistic map is to initial conditions in this regime. We find that it is not sensitive to initial conditions since the two lines follow each other almost identically and eventually converge to the same value.

We also evaluate the logistic map within the chaotic regime, as shown in Fig. 1b. This time, we use $r = 3.8$ and plot the initial population for $y_0 = 0.1$ and $y_0 = 0.101$ With this value of `r`, it becomes obvious that the logistic map is chaotic and extremely sensitive to the initial condition. Even though the difference in initial conditions is only 0.001, the population values start to become drastically different after roughly 11 time steps.

2) Entropy, Mutual Information: To further understand the previous data, we plot the entropy and mutual information for the beginning and end of each pair of plots. We define the "beginning" to be the first n timesteps, and the "end" to be the last n timesteps, where n is the number of timesteps before the chaotic series diverges. In this case, we have determined n to be 11. These plots can be seen in Fig. 2. To plot this data, we again used `Python3` and its `matplotlib` library, but we utilized the JIDT tool to perform calculations of entropy, mutual information, and transfer entropy [3].

To perform these computations, the data first needed to be discretized. In order to do this, we simply binned the values into 11 bins. That is, we put all values in `[0 - 0.1)` into a bin, then all values from `[0.1 - 0.2)` in a bin, etc. We used `Python`'s `numpy` library, and it's `digitize` function to accomplish this.

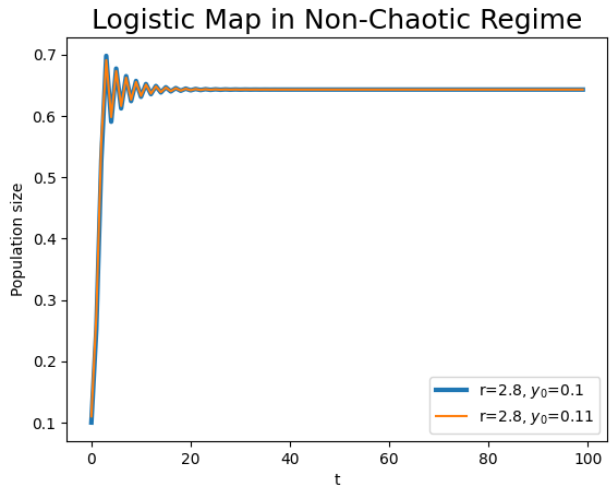
The first of these diagrams, Fig. 2a, has 0 entropy for each plot, and the mutual information takes up the entire diagram. This happens because the entropy of each line is exactly the same during this time range. Since the entropy is the same, the entire plot can be represented by mutual information.

The second diagram, Fig. 2b, is a totally empty plot. This is the result of the two lines converging to the same value, therefore sending both of the entropies to 0. When the entropies are 0, there is no possibility for mutual information, so the Venn Diagram stays completely empty.

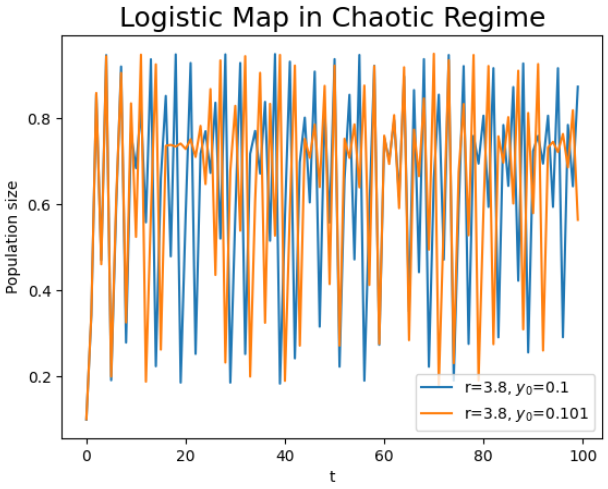
The third diagram, Fig. 2c, has a fair amount of mutual information that takes up a majority of the diagram. Intuitively this makes sense, since for the first 11 timesteps the two lines in the chaotic region follow each other pretty closely. They diverge a bit more than the beginning of the non-chaotic regime, but are close enough that there is a lot of mutual information.

The final diagram, Fig. 2d, shows the least (practical) amount of mutual information. At the end of the chaotic regime, the lines have significantly diverged, meaning the data is less similar and therefore has less mutual information.

Finally, we calculate the transfer entropy for the non-chaotic and chaotic regimes. The transfer entropy in the non-chaotic regime is simply 0.0. Since the two lines follow each other almost identically for the entire time, there is no information



(a) A plot of the logistic map in the non-chaotic regime, using an `r` value of 2.8 and initial populations of 0.1 and 0.11.



(b) A plot of the logistic map in the chaotic regime, using an `r` value of 3.8 and initial populations of 0.1 and 0.101.

Fig. 1: Plots of Non-chaotic and Chaotic Logistic Maps, respectively. The non-chaotic plot demonstrates a lack of sensitive dependence on initial conditions, while the chaotic plot demonstrates extremely sensitive dependence on initial conditions.

gain for one line by knowing previous values of the other. Therefore, it makes sense that the transfer entropy is `0`.

The transfer entropy for the chaotic regime is `0.49`. Since the two chaotic lines follow each other for a few timesteps, then begin to diverge, some amount of transfer entropy is expected. However, one line does not give sufficient information to completely predict the other, so it's still a relatively small amount of transfer entropy.

B. Bottom-Up vs. Top-Down

1) Exploring Transfer Entropy: To study the transfer entropy between two time-series, we again leveraged the JIDT [3] library. We set out to answer the question whether or not

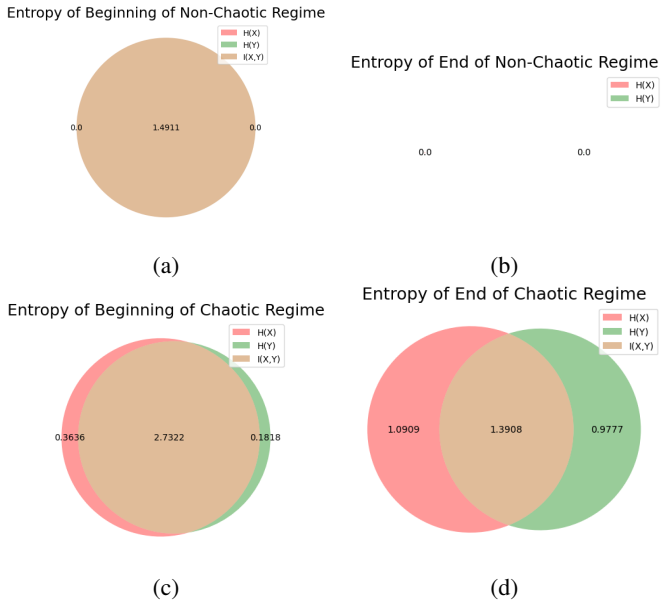


Fig. 2: Venn Diagrams showing the Entropy and Mutual Information for the first and last 11 timesteps of the non-chaotic and chaotic logistic map regimes. Fig. 2b is empty because the maps have converged to the same value, thus the entropies are 0 and there is no possibility for mutual information.

we can pinpoint a statistically significant switch from bottom-up to top-down causation for epsilon values ranging from 0.2 to 0.3. To test this theory, we replicated the toy logistic model from Walker et al. [7]:

$$x_{i,n+1} = (1-\epsilon)f_i(x_{i,n}) + \epsilon m_n; (i = 1, 2, ...N) \quad (3)$$

where the function $f_i(x_{i,n})$ specifies the local dynamics of element i, N is the total number of elements, n is the current time-step (generation), and ϵ is the global coupling strength to the instantaneous dynamics of the mean-field, m_n, defined below in eq. 5. The local dynamics of each element i is defined by the discrete logistic growth law:

$$f_i(x_{i,n}) = r_i x_{i,n}(1 - \frac{x_{i,n}}{K}) \quad (4)$$

where r_i is the reproductive fitness of population i, and K is the carrying capacity which is set to K = 100 for all i. Finally:

$$m_n = \frac{1}{N}\sum_{j=1}^{N} f_j(x_{j,n}) \quad (5)$$

is the mean-field which is a "global systemic entity" and cannot be tied back to any specific local attribute.

Simulations were run with parameters replicated from Walker as well, with K=100, N=1000, $x_{i,0} = 1$, and r_i drawn independently and uniformly from $[3.9, 4.0]$. Since we were specifically looking at epsilon values between 0.2 and 0.3, we only looked at the values $[0.2, 0.225, 0.25, 0.275, 0.3]$ for our simulations.

Once we had collected our simulated data, we calculated the transfer entropy from the instantaneous mean-field to the population data (top-down) and the population data to the instantaneous mean-field (bottom-up) for all generations for $k = 1, 2, 3, 4$ for each epsilon, where the max-k was used. We then calculated error bars to a 95% confidence interval. Figure 3 below shows the results of that analysis.

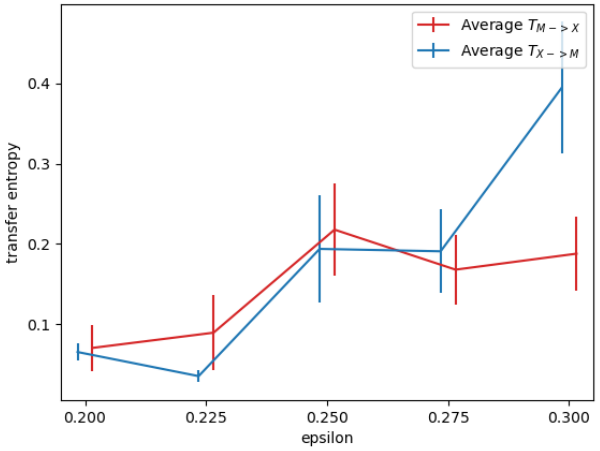


Fig. 3: Bottom-Up and Top-Down Transfer Entropy for values of epsilon between 0.2 and 0.3.

Here we can see clear overlap between the bottom-up and top-down values for each epsilon between 0.2 and 0.3 (exclusive) signifying that none of our values are statistically significant. Oddly enough, at epsilon 0.3 we see the bottom-up transfer entropy higher than the top-down, contradicting the result from Deterding and Wright.

2) Exploring Mutual Information: Next we looked at the average mutual information between the instantaneous mean-field and population data. Mutual information can be thought of as the shared information between two time-series. Again leveraging JIDT to perform the measurement over the Walker simulation data, we were able to produce Fig. 4.

In this figure, we see relatively large uncertainty (with respect to the values for transfer entropy) for some of the values further supporting the lack of statistical significance for an inflection point from bottom-up to top-down entropy.

C. COVID Data

1) Predicting Real-World COVID Data: In this section, we analyze COVID data in two U.S. states: New Mexico and Nevada. Our data is taken from `data.cdc.gov` [5]. We choose New Mexico because it is the authors' current state of residence. We choose Nevada because it is the state with the closest population density to New Mexico that is also a "blue" state (i.e. democratic) [2] [9]. Given that COVID regulations have been heavily politicized, we assume this means that the

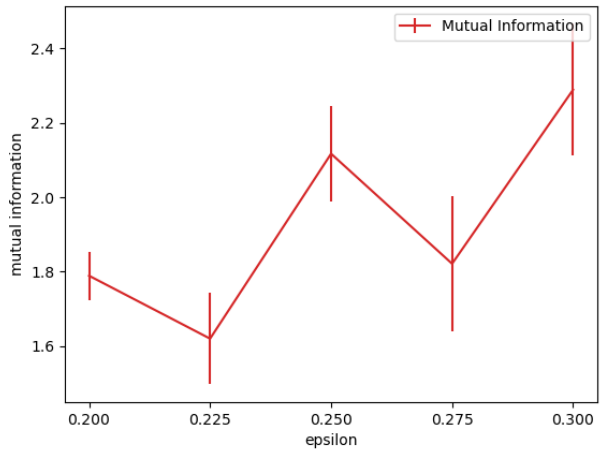


Fig. 4: Mututal Information between instantaneous mean-field and population sizes for values of epsilon between 0.2 and 0.3.

two states have had similar restrictions. We also choose to analyze data only from the start of the Omicron surge to the time of writing (`2021-12-01` — `2022-02-19`). Fig. 5 shows the number of new COVID cases for New Mexico and Nevada during this time range.

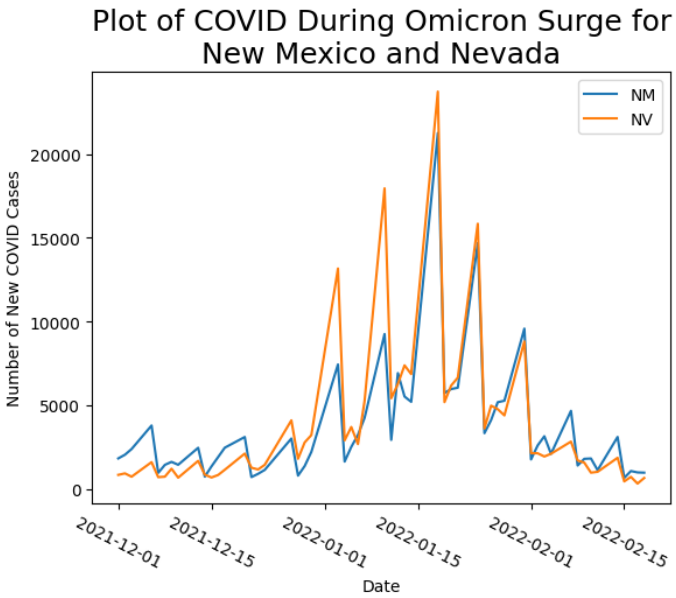
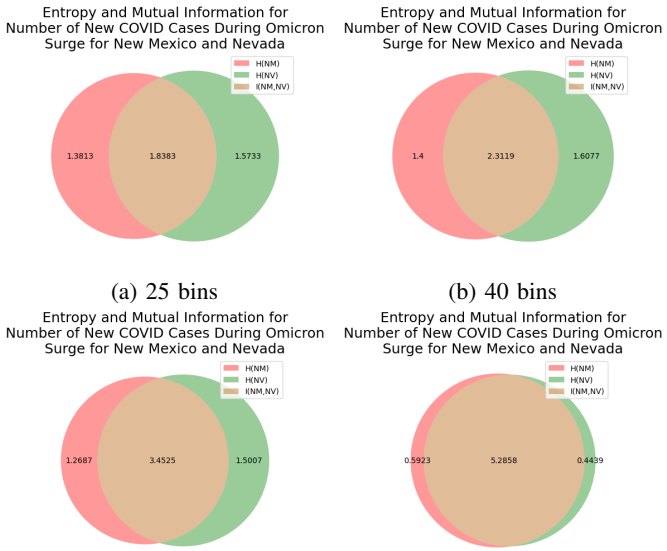


Fig. 5: A plot of number of new COVID cases during the Omicron surge for New Mexico and Nevada. The two lines share a nearly identical shape, but the scale differs somewhat.

Using this data, we try to determine if the dynamics of the Omicron variant in New Mexico can help predict the dynamics of it in Nevada, or vice versa. We experiment with a different number of bins for discretizing the data and a different `k` value (the number of timesteps to look backwards). The results can be seen in Table I. We notice that as we increase the number of bins and keep `k` constant, the transfer entropy in both directions increases. However, as we increase `k` and keep the number of bins constant, the transfer entropy decreases. Transfer entropy is at a maximum when `k` is `0`. This means that looking further back in time actually hinders the ability to predict the number of new cases in one state based on the other. This is likely because New Mexico and Nevada actually follow the same shape too closely. The historical data of one only provides minimal insight into the future value of the other since they are nearly the same distribution. This transfer entropy data indicates that there is little benefit in using New Mexico's or Nevada's data to predict the number of new cases in the other.

To further investigate this question, we use the mutual information metric. Fig. 6, shows the entropy and mutual information for this data with different number of bins used to discretize the data. We can clearly see that as the number of bins increases, the amount of mutual information increases substantially both in value and relative to the individual entropies. The plot with the highest number of bins (Fig. 6d) is closest to the real data because it will have lost the least amount of information by discretizing the data. In this plot, we see that there is significant mutual information and very small amounts of individual entropy. This also confirms the results that we saw from using the transfer entropy metric, that the two states' data is extremely similar.



(a) 25 bins (b) 40 bins

(c) 100 bins (d) 500 bins

Fig. 6: These Venn Diagrams show the mutual information for number of new COVID cases in New Mexico and Nevada during the Omicron surge. We see that as the number of bins increases, the relative amount of mutual information also grows. This helps to show us that the plots are quite similar, but without any transfer entropy, it is difficult to use one to predict the other.

Number of Bins	k	Transfer Entropy (NM → NV)	Transfer Entropy Ratio (NM → NV)	Transfer Entropy (NV → NM)	Transfer Entropy Ratio (NV → NM)
25	0	1.6374	1.6694	1.0407	1.2086
25	2	0.23807	0.4544	0.1513	0.3290
25	4	0.1079	0.1678	0.0686	0.1215
40	0	2.149	2.3341	1.3366	1.6672
40	2	0.2995	0.4364	0.1863	0.3117
40	4	0.0889	0.0789	0.0553	0.0564
100	0	3.5481	3.5427	2.3643	2.7923
100	2	0.1732	0.2148	0.1154	0.1693
100	4	n/a	n/a	n/a	n/a

TABLE I: This table shows the transfer entropy for various numbers of bins and k values between new COVID cases in New Mexico and Nevada during the Omicron surge. The columns that are "ratios" are the amount of transfer entropy from $A \rightarrow$B, divided by the entropy of B. This helps put the transfer entropy into perspective to provide more meaningful analysis.

If the dynamics of a particular COVID variant in one state could be used to predict the dynamics of that variant in another state, we would expect to see a different type of data. Specifically, we would expect a more significant amount of transfer entropy, especially when the k value was increased. We saw relatively high transfer entropy when k was 0, but transfer entropy decreased as k increased, meaning that prediction would be difficult. We do, however, see a large amount of mutual information, as would be expected if that statement were true. In order to predict one state based on another, there should be at least a fair amount of mutual information. We did observe this, but with the lack of transfer entropy, any prediction techniques would likely not be very successful.

2) COVID Data in Southwestern United States: Finally, we extend our analysis from II-C1 to look at bordering states in the southwestern US. Given that inter-state migrations are almost always occurring between populations in some capacity, this is a natural extension from the local Nevada to New Mexico, and vice versa, scope. To supplement our existing analysis, we added Texas, Colorado, Utah, and Arizona to our list of states to calculate bidirectional transfer entropy. We also broke down each combination by variant, adding alpha and delta to the analysis in figure 7.

As discussed in Wu et al., a classical logistic model is one of the less flexible growth models due to its inability to estimate the final capacity of the population. This happens because it fails to account for the super exponential growth and the potential slow abating of the epidemic [10]. Despite these limitations, we attempted to fit the logistic model to the average new case count by variant across all states. Our approach was to smooth the data and optimize for the r-value that minimized the mean-squared error. This approach though was problematic and we were unable to come to any meaningful results, so instead we measured the transfer entropy across all states and filtered by variant. The results of that graph are in the Fig.8.

Although the alpha variant had the lowest transfer entropy, this could be related to the time phasing of dates by variant. As noted in Section II-C1, Omicron was classified as any new cases after 12/1/2021, Delta was any new case between 5/1/2021 and 12/1/2021, and Alpha was any new case before

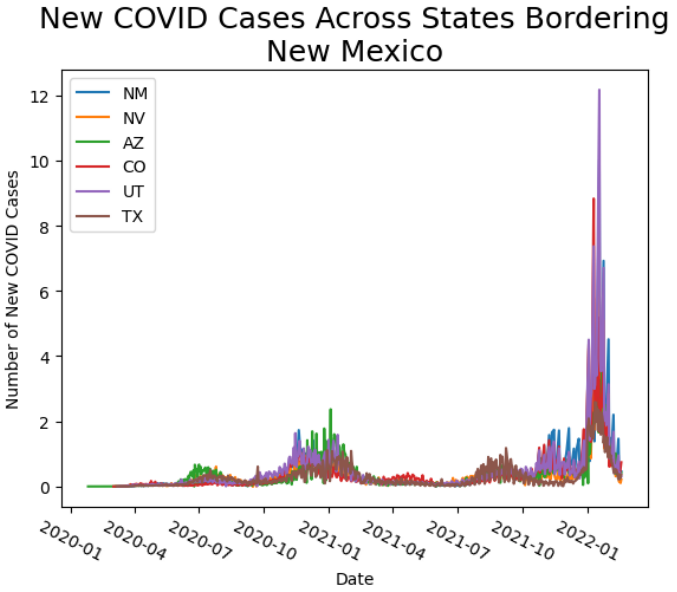


Fig. 7: Average new cases per capita in Southwestern United States

5/1/2021. This leaves a much longer time period for alpha, despite being normalized per capita, which could be negatively impacting the transfer entropy calculation.

Generally, a transfer entropy of zero means that the time series in question were in complete synchronization since no information can be gained by considering the history of a second time series. Extending this analysis to our results, this implies that the Alpha COVID-19 variant performed characteristically similar between all areas around the same time and Delta and Omicron had a lag in their spread leading to the opportunity for this information gain. In perhaps a less rigorous analysis and closer to an observation, this shape of transfer entropy between the variants mirrors the transmissibility of virus over time.

III. Discussion/Conclusion

In the Walker paper, transfer entropy was measured between various populations and a mean-field, where the former represented the "bottom" part of a causal structure and the

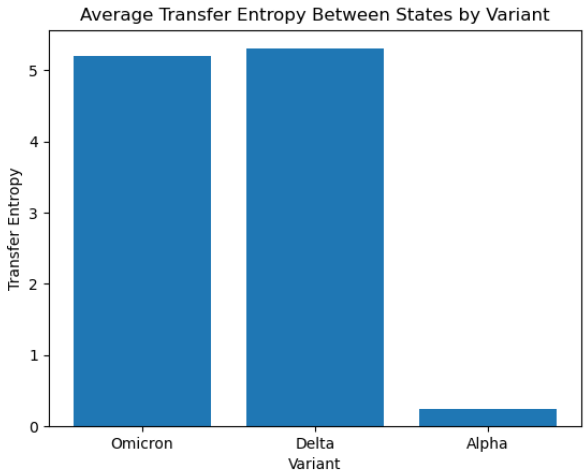


Fig. 8: Plot of the average transfer entropy for the best fit logistic model across average new cases per capita between NM, AZ, NV, CO, UT, and TX.

latter the "top." Their exploration with the toy logistic map model concluded that there was a perceptible difference between bottom-up and top-down transfer entropy. Deterding and Wright confirmed this, though raised the point that it's unclear where the switch from bottom-up to top-down happens as the influence of the instantaneous mean-field grows. Our analysis yielded similar, inconclusive results.

From our results, we can determine that the logistic map and related tools are moderately effective for performing analysis of COVID data. We find that the states that that we analyzed all had very similar shapes, and thus high mutual information. The transfer entropy for the different states was high during the Omicron and Delta variants, as shown in Fig. 8, which implies that transfer entropy can be used to help predict trends between states during these variants. The Alpha variant was the main outlier in terms of transfer entropy. The transfer entropy was quite low during this time, which may be a result of the length of time that the variant persisted. However, it could also imply that the Alpha variant was so sporadic that there was actually very little transfer entropy and possibility for prediction. As a result, we conclude that transfer entropy can be a reasonable indicator that prediction is possible, but it should be verified with other methods. Future work could include the use transfer entropy to take a look at how transmissibility rates change over time given variant evolution and geographic location.

From our analysis, we determine that top-down drivers seem to affect COVID-19 dynamics more than bottom-up. We consider our study of mutual information and transfer entropy to be top-down because we track states with similar mandates or similar location. We consider the bottom-up approach to be the logistic map. When trying to fit the logistic map to these plots, we were largely unsuccessful and found nothing that sufficiently represented the data. However, using the top-down approaches provided better insight, with some of our transfer entropy results indicating some promise in prediction techniques. These results imply that mandates and human behavior are affecting the spread of COVID-19 more than the natural contagiousness of the disease, but we leave it as future work to verify these results using other methods.

IV. Contributions

The authors of this team contributed well to complete this project. Greg handled all of part 1 (code, figures, writing that section of the report) while Colin handled the same things for part 2. We split up part 3 with Greg handling the code and report for the first half, and Colin expanded on that work to handle the rest of the analysis. We shared the other sections of writing the report based on whose portion of the project that section was more closely aligned with. Although we mostly maintained the above distribution of work, there was a fair amount collaboration for debugging errors and sharing ideas throughout the duration of the project.

All of the code we used in this paper can be found at `https://github.com/gregjacobus-school/unm-cs523-project1`.

References

[1] Justin Detering and Catherine Wright. The coupled logistic map, 2020.
[2] GKGIGS. List of blue states and red states in u.s. 2022 (updated). https://www.gkgigs.com/list-of-blue-states-and-red-states/.
[3] jlizier. Jidt. https://github.com/jlizier/jidt.
[4] Melanie Mitchell. Complex adaptive systems.
[5] Surveillance Review and Response Group. United states covid-19 cases and deaths by state over time. https://www.gkgigs.com/list-of-blue-states-and-red-states/.
[6] T Schreiber. Measuring information transfer. phys. rev. lett., 85:461., 2007.
[7] Sara Imari Walker, Luis Cisneros, and Paul C. W. Davies. Evolutionary transitions and top-down causation. arxiv:1207.4808v1 [nlin.ao, 2012.
[8] wdjpng. Chaos - the logistic map. https://colab.research.google.com/github/wdjpng/chaos/blob/main/main.ipynb.
[9] Wikipedia. List of states and territories of the united states by population density. https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States_by_population_density.
[10] Ke Wu, Didier Darcet, Qian Wang, and Didier Sornette. Generalized logistic growth modeling of the COVID-19 outbreak: comparing the dynamics in the 29 provinces in china and in the rest of the world. *Nonlinear Dynamics*, 101(3):1561–1581, August 2020.