What can the Logistic Map explain about COVID dynamics?

1st Liangkun Yu
Dept. of Electrical and Computer Engineering
University of New Mexico)
liangkun@unm.edu

2nd Chaeeun Park
Dept. of Computer Science
University of New Mexico)
chaeeun0618@unm.edu

Abstract—We replicate Walker's paper for the two-way, top-down, and bottom-up, vertical causation analysis in biological systems with the globally coupled logistic map. We can only find consistent significant transfer entropy in the hierarchical system for some specific interval, $\epsilon > 0.5$. We apply both vertical and horizontal causation ideas to study the Covid-19 cases in the United States and draw some interesting conclusions regarding early-stage phenomenon and vaccination. But for the real conclusion about the matching between logistic map and COVID. We found some similarities, while more difficulties for applying logistic map to COVID as the result of lacking self-consistent design.

Index Terms—component, formatting, style, styling, insert

I. INTRODUCTION

The logistic map is a function chosen for replicating the biology population growth pattern in [1]. Although logistic map is a simple function generating a sequence of deterministic data, it still could emerge chaotic behaviors.

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

where x_t is the population for the species at time t and the legal value for growth rate R is $0 \le R \le 4$. The specific range for R and the corresponding behavior can be found in Table.I. Except for a few values (like $R = 1 + \sqrt{8}$), most values of R beyond 3.56995 exhibit chaotic behaviour.

Actually, logistic map is not only sensitive to

In the long term, interpretation of the relation between nature and biology species is dominated by Darwin's theory that is natural selection. Only the individuals with high fitness to the surrounding environments could survive. This concept focuses on explaining the mechanism sin biological organization with a top-down view. Only the top can influence, even determine the biology species. But Walk believes that there is a one-way impact that biological systems and nature must exist some kind of interactions [2]. Therefore, when we try to explain the mechanism in the biological systems, we must also consider the bottom-up effect.

To under the interaction between top-down and bottom-up, we can use the globally coupled logistic map as to toy model to simulate the population growth for different species $\mathcal{I}$ and the interaction among them.

$$x_{i,t+1} = (1 - \epsilon)R_i x_{i,t}(1 - x_{i,t}) + \epsilon m_t, \tag{2}$$

where i is one species among all available species $\mathcal{I}$, ϵ is defined as global coupling coefficient, and m_t is the average over all species' population at time t that,

$$m_t = \sum_{j=1}^{|\mathcal{I}|} x_{j,t}. \tag{3}$$

Walker intends to use logistic map function to reflect the interactions between a specific and the higher level (which could be considered as environment). But actually, it is not a good function to simulate the COVID-19 infection as it failed to consider the direct horizontal causation. There is no connection that can be established between species A and species B for using logistic map even with the global coupling coefficient. We still analysis both vertical and horizontal causation for COVID cases in US.

TABLE I: Logistic map patterns for different R

R	Pattern
$R = 0$	Doom to extinction
$R \le 1$	Doom to extinction
$1 < R < 3$	Converge to fixed point
$3 \le (1 + \sqrt{6})$	Period-2 cycle
$3.44949 < R < 3.54409$	Period-4 cycle
$\vdots$	Higher order period cycles
$3.56995 < R < 4$	Chaotic behavior
$4 = R$	Increase at the maximum rate
$4 < R$	Leave the interval and diverge

II. METHODS & RESULTS

A. Basic concepts

We begin with reviewing Shannon entropy, mutual information, and transfer entropy.

1) Shannon Entropy (Entropy): Each time sequence could contain some information, which is measured by Shannon entropy. Actually, the entropy method quantifies the amount of regularity and the unpredictability of fluctuations over time-series data [3]. The exact method is to quantify the continuous integer population into different bins and measure the "uncertainty" and probability for each bin,

$$H(X) = -\sum_{i=1}^{B} p(x_b)log_2(p(x_b)), \tag{4}$$

TABLE II: Summaries for MI and TE

	Similarity	Difference
MI	measured in bits	sequence invariant
TE	between two random processes	sequence sensitive

where $p(x_b)$ is the probability of time series samples appearing in the b-th bin, and the "uncertainty" of a bin is measured with $log_2(p(x_b))$, which reflects the Shannon information in bits. Besides, the number of bins B is determined with Sturge's rule [4],

$$B = 1 + ceil(\log_2(\beta)), \quad (5)$$

where β is the sample size. In our report, $\beta = 10$.

2) Mutual Information (MI): Now we have two-time series (X and Y), and two Shannon entropy ($H(X)$ and $H(Y)$). If the two series are independent, then there is no mutual information. But if there could be some potential factors link the two-time series, then mutual information measures how much one random variable tells us about another, which is measured with,

$$I(X,Y) = \sum_{x \in X} \sum_{y \in Y} P_{XY}(x,y) log2 \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}. \quad (6)$$

3) Transfer Entropy (TE): Transfer entropy is similar to mutual information measuring the direct transfer of information between two random processes. But the difference is that transfer entropy focus on the sequence time, while mutual information does not care about the sequence, summarized in Table.II. Transfer entropy from a process X to another process Y is the amount of uncertainty reduced in future values of Y by knowing the past values of X given past values of Y [5].

$$T_{X \to Y} = H(Y_t|Y_{t-1:t-L}) - H(Y_t|Y_{t-1:t-L}, X_{t-1:t-L}) \quad (7)$$
$$= I(Y_t; X_{t-1:t-L}|Y_{t-1:t-L}) \quad (8)$$

where L is the time depth for the future prediction based on history.

B. Part 1

We use logistic map to simulate the population growth for two individual and independent groups. The logistic map is slight from Eq.1 by incorporating carrying capacity,

$$x_{t+1} = Rx_t(1 - \frac{x_t}{K}), \quad (9)$$

In Fig.1, there are two graphs corresponding to two times series with different growth rates, which are ($R = 2.0$) for the top set and ($R = 3.9$) for the bottom set. In each set, there are two times series corresponding to different initial populations. And the initial population setting is the same across different sets. The carrying capacity $K = 100$ and the overall time length is also 100. Also, we can observe from the graphs, the top set having a non-chaotic growth rate ($R = 2.0$) soon converges to 50 population around 9-th round. But for the bottom set with chaotic growth rate ($R = 3.9$), they never

converge to a specific value and there is no pattern that could be found from the oscillation. The observations from Fig.1 match the growth rate pattern summation shown in Table.I.
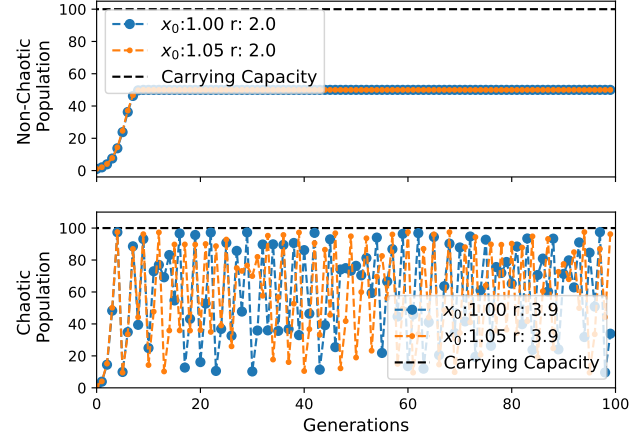

Fig. 1: Time series of a non-chaotic population (top set) and chaotic population (bottom set) Reproductive coefficients (2.0,3.9) are equal for each series in a set. Initial populations are varied in each set (1.00,1.05).

We replicate code from [6] for the Venn diagrams in 2. We keep 10 bins for measuring the entropy. The MI is generated over Discrete algorithm of base 10 in JIDT.

For the first graph [1,1], the only difference two series are the initial population. Two series quickly converge to the population of 50. They basically overlap with each other. So the MI in the first 10 time slots is also overlapped, $I(X,Y) = 2.05$. For the second graph [1,2], the high growth rate $R = 3.9$ amplifies the difference in the initial population. So there, two series are not consistent in the first 10. Visually, MI is smaller compared to the first graph. But actually, the dynamic range is bigger than the MI is bigger, $I(X,Y) = 2.17$. For the third graph [2,1], it's corresponding to the last 10 time samples in Fig.1 top set. They never deviate from the converged value. So the entropy for them is 0. Also, the MI $I(X,Y) = 0$. For the fourth graph [2,2], with amplification of high R, two time series have deviated from each other. But there is still some trace between them, that MI $I(X,Y) = 2.05$.

C. Part 2

In this part, we generate new ϵ interval for TE and MI calculation. We get 24 points from each intervals, $[0.2, 0.3)$ and $[0.3, 0.7]$. Here, the global coupled logistic map is used to generation the population data,

$$x_{i,t+1} = (1-\epsilon)R_i x_{i,t}(1 - \frac{x_{i,t}}{K}) + \epsilon m_t. \quad (10)$$

The related data for simulation can be found in Table.III. We repeat 10 times for generating confidence interval. Among 1000 sub-groups, 3 sub-groups are randomly picked for further calculation. We follow the same path to calculated MI shown
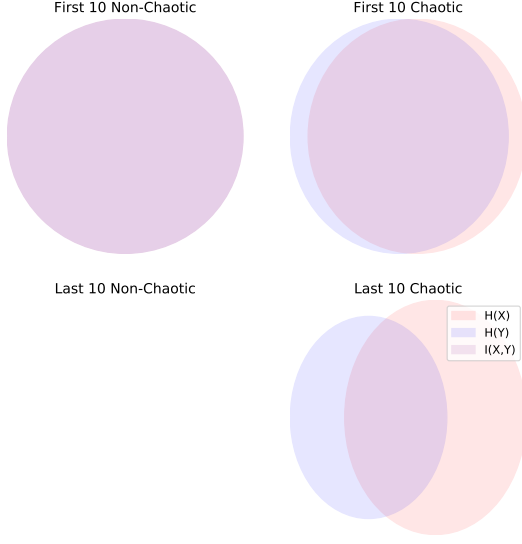
Fig. 2: Top set corresponds to the first 10 samples. Bottom set corresponds to the last 10 samples. Left set corresponds to non-chaotic $R = 2.0$. Right set corresponds to chaotic $R = 3.9$. The MI for them are $[2.05, 2.17, 0.00, 2.05]$.

TABLE III: Global coupled logistic map parameters

Logistic map parameter	Value
Number of repeating time	10
Number of individual groups ($|\mathcal{I}|$)	1000
Number of generations (T)	1000
Growth rate (R)	Ramdomly pick (3.9, 4.0)
Carrying capacity (K)	1000
Initial population (x_0)	1

Fig. 3: Average mutual information (MI) measured in bits between sub-groups 95% confidence interval of MI.

in Fig.3 and TE shown in Fig.4. In Fig.3, mutual information grows linearly to the increase of ϵ among $[0.2, 0.7]$, and the confidence interval becomes more concentrated.

1) TE calculated over JIDT: TE is generated by tracing up to 2 time depth into the past, as the constrain from JAVA heap limitation. But the data in Fig.4 looks specious. There is no confidence interval for the data generated in Walker's paper. Because Walker already takes the maximum TE value from the available list that

$$T_{X \to Y} = max \left\{ T_{X \to Y}^k \right\}. \tag{11}$$

where k belongs to all available data points for TE calculation.

But in Justin's report, he keeps the maximum TE. And based on stand deviation of the maximum TE list to calculate the TE confidence interval. I believe the right way to do it is to keep the average TE,

$$T_{X \to Y} = average \left\{ T_{X \to Y}^k \right\}, \tag{12}$$

and calculate stand deviation based on averaged TE. Then, we can get the right confidence interval.
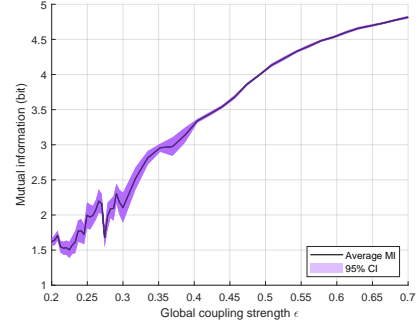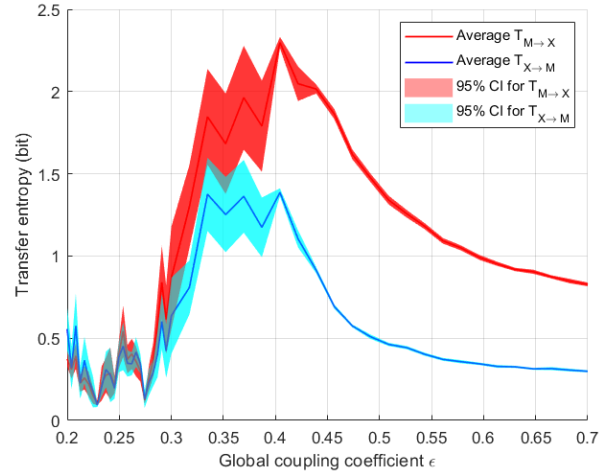

Fig. 4: Maximum top-down ($T_{M \to X}$) and bottom-up ($T_{X \to M}$) transfer entropy (calculated over JIDT with $k \leq 2$). This graph is problematic for the 95% confidence interval. Confidence interval should be calculated over average value rather than maximum.

Due to two reasons, 1) JIDT can not calculate depth beyond 2; 2) The concept errors in Justin's realization of

TE confidence interval, We decide to use the transfer entropy package in R language.

2) TE calculated over R: We only use one set of randomly generated datasets for the mean-field (M) with one sub-groups (X) for each ϵ. The search space is

$$48 \times 3 \times 4. \tag{13}$$

They are corresponding to sample points for ϵ interval, 3 randomly picked sub-groups and 4 time depth.

We do not draw the confidence interval, but bootstrap is set to 300, which is big enough to generate a p-value. Bootstrap retains the uni-variate dependencies for M but eliminates the statistical dependencies between M and X [7]. P-value indicates the significance of the observation, usually with $p - value < 0.05$. Therefore, we keep the maximum transfer entropy, but only keep the TE with statistic significance shown in Fig.5.

$$T_{X \to Y} = \begin{cases} max \left\{ T_{X \to Y}^k \right\}, & pvalue^k < 0.05, \\ 0, & otherwise, \end{cases} \tag{14}$$

There are some similarities between Fig.4 and Fig.5. For the top-down transfer entropy between ϵ interval $(0.2, 0.45)$, the 95% confidence interval almost overlap with each other, making the transfer entropy not significant shown in Fig.4. which corresponds to the inconsistency of the significance during the same interval in Fig.5. The top-down TE oscillates a lot.

Actually, in Fig.5, top-down transfer entropy only is significant for the interval $(0.5, 0.7)$. And bottom-up transfer entropy only is significant around 0.5.

Besides, R using Kullback-Leibler distance to measure the entropy makes the value scale much smaller, like all entropy $< 0.1 bits$ in Fig.5.

Combining both factors that 1) limited range of consistent significance and 2) small scale of transfer entropy, it's hard for us to find the interactions between top-bottom in the interval of $\epsilon \in [0.2, 0.7]$.

D. Part 3

In Walker's model, each sub-group only has one connection, which is a two-way connection to the higher level shown in Fig.6. Here, we want to study COVID cases "$cases\_avg\_per\_100k$" and we follow Walker's idea assuming the interactions only exist between US (country level) and (state level). California, as the most populous state in the US, is picked for study.

First, based solely on the view of bottom-up causation, the increase in the case of the NM also increases the cases in the US. To combat COVID, a series of measures are proposed, varied by the subject of the promulgation, the object of the implementation, and the time. To attribute all measures over the idea of the direction of causation, we may label vaccination, stimulus check as the top-down and lab travel restrictions, population, age structure as the bottom-up. As a consequence, two-way causation will pass the upward causation from NM over the top-down channel to other states.
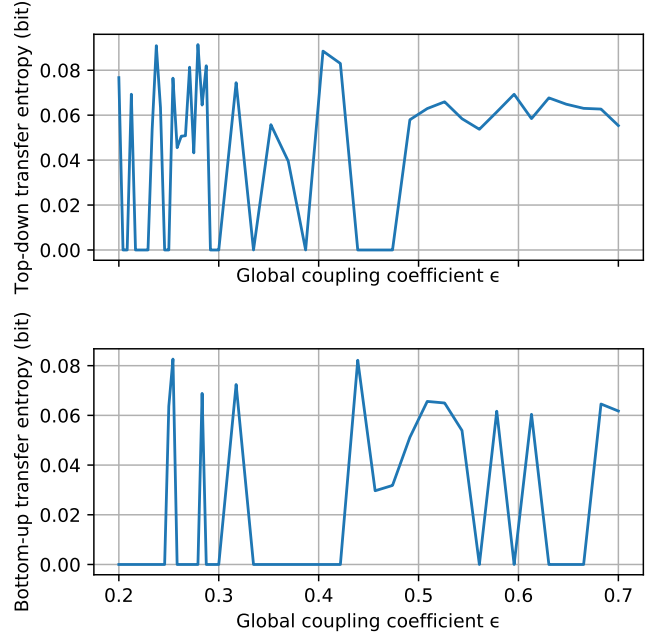

Fig. 5: Transfer entropy (calculated over R with $k \leq 4$) between mean-field and one subgroup. Top set for top-down causation and bottom set for bottom-up causation. Only the TE with p-value significance are non-zero. Only the top-down $0.5 < \epsilon < 0.7$ and bottom ϵ around 0.5 have persistent significant transfer entropy.
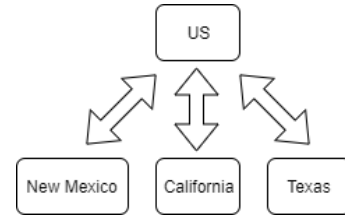

Fig. 6: Vertical causation. No state level direct horizontal causation.

This means increased cases in NM will increase the probability of cases in other states over the interactions to the US.

The value of mutual information between US and CA is large. It also has a significant transfer entropy from US to CA. It can be said that this is representing the top-down causal relationship between them. We divided into Before vaccine massive administration and After vaccine massive administration. As can be seen in Fig.7, the mutual information between After vacuum passive administration US and CA increases.

$$I(US, CA) = 1.232 \tag{15}$$
$$> H(US|CA) \tag{16}$$
$$> H(CA|US). \tag{17}$$

After the vaccine administration, there is still a significant transfer entropy, but it has not increased. There are many
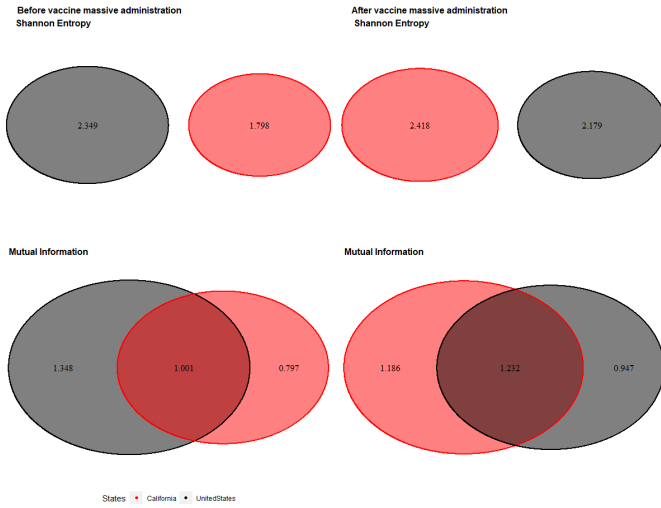
Fig. 7: Venn diagrams illustrating the relative scale of the entropy, and the mutual information measures for the cases of US and CA.
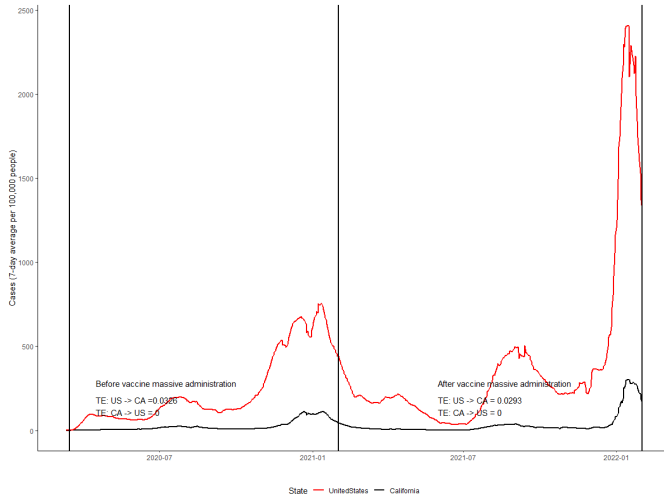

Fig. 8: Transfer entropy between US and CA during [03-15-2020,02-01-2021], and [02-01-2021,2-15-2022]. There is significant transfer entropy from US to CA during all intervals.

variables there. One of the reasons for the significant change in January 2022 is the emergence of the Omicron mutation virus and its tremendous propagation power.

Our hypothesis is "The dynamics of a particular COVID variant in one country (or region) predict the later dynamics of that variant in other countries (or regions)." In the US and CA, vaccination and policy intervention are almost the same because CA is part of US. The relationship between US and CA was a Top-down relationship, so mutual information and transfer entropy showed the results of proving the hypothesis as we analyzed above.

Actually, the transfer entropy result matches our finding in Fig.5. In the US-CA case, only US has persistent significant transfer entropy to CA. Just like the phenomenon in Fig.5,

the top-down transfer entropy is also persistent significant. Although they have similar phenomena, we can still not able to conclude that the global logistic map is a good model for covid.

E. Part 4

We investigate the mutual information and transfer entropy between mean-field (M) and a sub-group (X) in the last part. The global coupled logistic map is used to reveal the bottom-up causation. But the fact is the states have direct horizontal causation among states without the detouring from state A to the county level to state B. Therefore, in this part, we investigate the mutual information and transfer entropy among states for two topics, 1) early-stage, 2) vaccination.
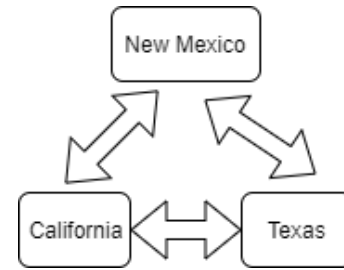

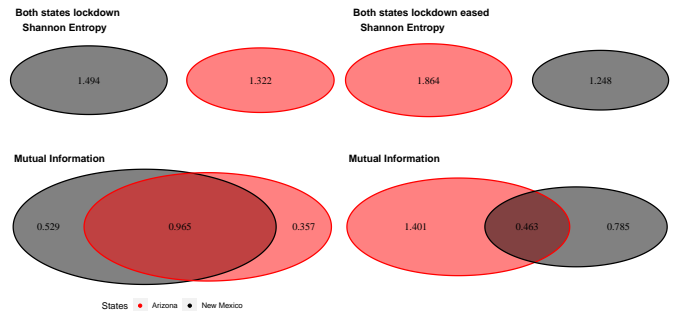Fig. 9: Horizontal causation. Only state-level horizontal direct causation.

Fig. 10: Mutual information between NM and AZ during [03-15-2020,05-1-2020], and [05-1-2020,09-15-2020]. Mutual information is high during the early-stage lockdown, and decreases a lot after lockdown lifted.

1) Early-stage: We measure the relation between NM and AZ at the beginning period from 03-01-2020 to 09-01-2020. This period is roughly divided into two sub-periods at 05-15-2020. In the first sub-period, both states execute different levels of lockdown. Arizona locks down the whole state from the middle of March to the middle of May, and New Mexico locks down Gallup for the first two of May.

During the lockdown period, there is not much traffic (people flow) in/out of states. So we can consider the two states as two independent processes. Also, it's the beginning period for both states. Mutual information for NM and AZ during the
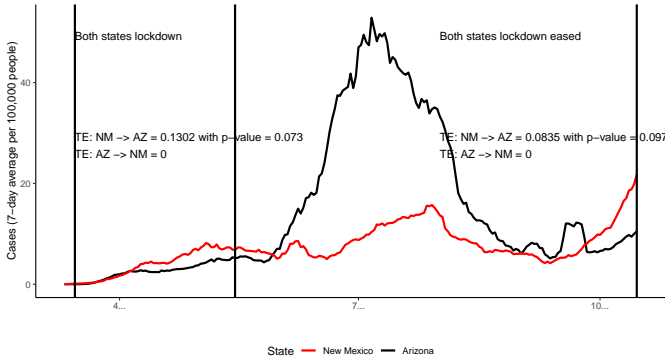
Fig. 11: Transfer entropy between NM and AZ during [03-15-2020,05-1-2020], and [05-1-2020,09-15-2020]. There is no significant transfer entropy between them, none p-value less than 0.05.

lockdown period is shown in Fig.10. The mutual information is greater than the conditioned entropy, that

$$I(NM, AZ) = 0.965 \tag{18}$$
$$> H(NM|AZ) \tag{19}$$
$$> H(AZ|NM). \tag{20}$$

But the high mutual information cannot be used to illustrate the high connection between two adjacent states, NM and AZ. The only significant reason to explain the high mutual information is that both states are at the very beginning. Just like the mutual information for the first-10 time-series samples in Fig.3. Covid-19 exhibits a similar growth pattern in both states. At the same time, the insignificant transfer entropy shown in Fig.11 can also support the above statement.

After 05-15-2020, both states lift the lockdown. Then two states deviate from each other by having much less mutual information $I(NM, AZ) = 0.463$. I believe the weakened mutual information can be explained for two reasons. 1) Small deviation during the process leads to huge discrepancy like the last-10 samples' mutual information shown in Fig.3; 2) With eased lockdown, the external factors take dominance. NM and AZ are states with a small population. NM is easy to be affected by TX, correspondingly, AZ is easy to be affected by CA.

2) Vaccination: We investigate the impact of vaccination to Covid-19 shown in Fig.12 and Fig.13. We pick Vermont as the highest fully vaccinated state and Wyoming as the lowest vaccinated state. Mutual information increases after vaccine administration.

As we can see from Fig.12, that after the vaccine administration. The mutual information between VM and WY increased, from 0.561 to 0.955. Besides, in the transfer entropy part in Fig.13, it has significant two-way transfer entropy after the massive vaccine administration.

Actually, even combining the plots, it's hard to draw any conclusion regarding the vaccine effectiveness. Vermont has a steady low case for a long time but also has a higher peak than
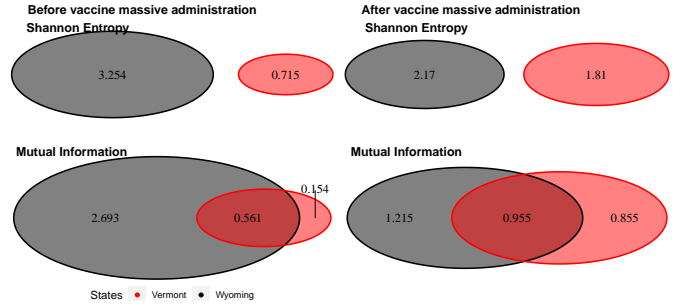

Fig. 12: Mutual information between NM and AZ during [03-15-2020,02-01-2021], and [02-01-2021,2-15-2022].
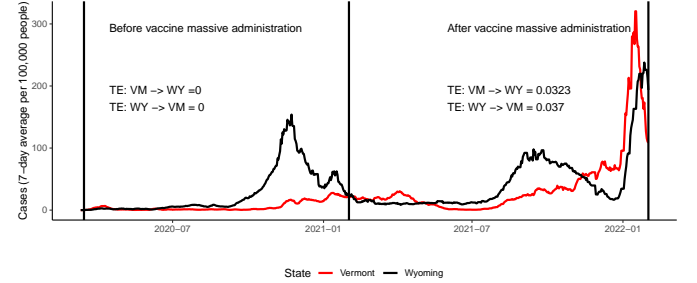

Fig. 13: Transfer entropy between VM and WY during [03-15-2020,02-01-2021], and [02-01-2021,2-15-2022]. There is a significant two-way transfer entropy after the vaccine administration as more restrictions are lifted during [02-01-2021,2-15-2022].

Wyoming. Integrating the news, we know that with the fully vaccinated percentage increases, more restrictions are eased and US life restored to old days. With more traveling people, even VM and WY, two far-away states, could have significant transfer entropy to each other.

III. CONCLUSIONS

We found some problems in Walker's paper. He use the globally coupled logistic map to investigate the vertical two-way causation between genes (bottom) and environments (top). He states there is observable transfer entropy in the two-way direction when $\epsilon > 0.3$. But we only observe the persistent and significant transfer entropy happening for the top-down during $0.5 < \epsilon < 0.7$. Our observation goes against some part of his conclusions.

Basically, we can consider the bottom level as species and there is no direct horizontal causation in genes level. If we assume the species as competitors or (preys and predators) in the system, then the value of ϵ should be negative. It seems the only fit scenario is assuming the species are mutually beneficial. Then we could have a positive ϵ. But if it only fit the mutual benefit scenario, how can we use the logistic map to mimic the information flow in biological systems.

Is a logistic map is a good model for COVID? Yes and No. We observe a high similarity between Fig.2 and Fig.11. Here, the conclusion is that the logistic map is a good function to

mimic natural growth, especially at the early stage. But there are too many factors, like eased restriction and affection from neighboring bigger states, that stop us from getting further conclusions. Another part of the supporting research from the US-CA case. We believe the ϵ should be small in the early stage and big in the later period. Because there are more restrictions in the early stage. Although more vaccines are administrated, more restrictions are also eased at the same time, making ϵ goes higher. Therefore, only the second part of US-CA case exhibits the significant transfer entropy.

While part of the answer is also no. Resort to the design of the global coupled logistic map, there are no direct interactions among species at the bottom. Species A must first affect the mean-field to affect another species. But in the Covid case, a state can directly affect other states, especially the adjacent states.

CONTRIBUTION STATEMENT

We form a supergroup of eight persons to finish this project. The main contribution from Eliza Gilbert. She guided me through using R language and explaining a lot of statistics and complex systems concepts to me.

We are a team of two. Within our group, Chaeeun generates Fig.1, Fig.2, Fig.7, and Fig.8. Also, she is responsible for the writing in Introduction and Part 3. Liangkun does the rest.

REFERENCES

[1] J. D. Murray, *Mathematical biology II: spatial models and biomedical applications*. Springer New York, 2001, vol. 3.

[2] S. I. Walker, L. Cisneros, and P. C. Davies, "Evolutionary transitions and top-down causation," *arXiv preprint arXiv:1207.4808*, 2012.

[3] S. M. Pincus, I. M. Gladstone, and R. A. Ehrenkranz, "A regularity statistic for medical data analysis," *Journal of clinical monitoring*, vol. 7, no. 4, pp. 335–345, 1991.

[4] D. W. Scott, "Sturges' rule," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 1, no. 3, pp. 303–306, 2009.

[5] M. Wibral, R. Vicente, and J. T. Lizier, *Directed information measures in neuroscience*. Springer, 2014, vol. 724.

[6] C. W. Justin Deterding. (2020) The coupled logistic map. [Online]. Available: https://github.com/judeter/CS523-TopDownCausation/blob/master/CS523_Project_1.pdf

[7] T. Dimpfl and F. J. Peter, "Using transfer entropy to measure information flows between financial markets," *Studies in Nonlinear Dynamics and Econometrics*, vol. 17, no. 1, pp. 85–102, 2013.