Analyzing Effectiveness of Gang Interventions using Koopman Operator Theory

Sian Wen, Andy Chen, Tanishq Bhatia,
Nicholas Liskij, David Hyde, Andrea L. Bertozzi
Department of Mathematics
UCLA
Los Angeles, CA
sianwen22@g.ucla.edu, andyxc@g.ucla.edu, tanishqbh@ucla.edu,
nliskij@ucla.edu, dabh@math.ucla.edu, bertozzi@math.ucla.edu

P. Jeffrey Brantingham
Department of Anthropology
UCLA
Los Angeles, CA
branting@ucla.edu

***Abstract*—Koopman operator theory, applied via numerical techniques such as dynamic mode decomposition (DMD) and autoencoders, has recently emerged as an interesting mathematical framework for understanding how complex, high-dimensional dynamical systems evolve. In this paper, we apply several DMD and autoencoder algorithms to a dataset of gang involvement and activity to assess the effectiveness City of Los Angeles Mayor's Office of Gang Reduction and Youth Development's (GRYD) Intervention Family Case Management Program. We compare various subsets of the data to explore differences in sub-populations. We then control for different covariates in our analysis of dynamical changes in population characteristics over time. Statistically significant results suggest the efficacy of the GRYD FCM Program.**

***Index Terms*—dynamic mode decomposition, gang membership, gang activity, autoencoder**

I. Introduction

Various institutional, non-profit, and government groups have introduced programs whose goal is to ameliorate gang crime and activity in their communities [1]–[5]. In particular, the City of Los Angeles Mayor's Office of Gang Reduction & Youth Development (GRYD) was established in 2008 to fund and oversee gang prevention and intervention services in various areas across the city [6]. GRYD Prevention services are aimed at building resilience to risk factors among at-risk youth (ages 10–15) to reduce the likelihood that they join gangs. GRYD Intervention Family Case Management (FCM) services are aimed at increasing youth and family protective factors and resiliency while also reducing gang embeddedness for gang-involved youth and young adults (ages 14-25). Although eligibility for intervention services is not determined using an eligibility tool [6], referrals to the program complete the Social Embeddedness Tool (SET) questionnaire. The SET questionnaire is intended to assess gang involvement and activity, but it also asks about other important covariates such as lifestyle, family life, gender, age, etc. Additionally, GRYD FCM Providers record the problems and strategies identified by participants and the FCM Strategy Team when building case plans. Both the SET data and the additional data on participant problems and strategies to address the problem were used in this analysis. Some who participate in GRYD FCM services complete the SET questionnaire multiple times, yielding insight into the effectiveness of GRYD FCM services in reducing individual gang involvement and activity.

We model SET responses at different points in time as sample points on the manifold of a highly non-linear dynamical system whose time evolution operator is unknown. Notationally, we let $\mathcal{M} \subset \mathbb{R}^m$ be a dynamical system on a finite-dimensional manifold which evolves according to the equation

$$z_{k+1} = \varphi(z_k), \tag{1}$$

where z_k represents the state of a system at time k and $\varphi : \mathcal{M} \to \mathcal{M}$ is a map that advances the system one step forward in time. Then we can define the Koopman operator $\mathcal{K}_\varphi$ which acts on scalar functions $f : \mathcal{M} \to \mathbb{R}$ such that

$$\mathcal{K}_\varphi f(z) = f \circ \varphi(z) \tag{2}$$

for all $z \in \mathcal{M}$ [7]–[10].

Given a set of scalar observations of the system f_k, predicting the same set of observations one step forward in time is equivalent to applying the Koopman operator to f_k, i.e.

$$f_{k+1} = f(z_{k+1}) = f \circ \varphi(z_k) = \mathcal{K}_\varphi f(z_k) = \mathcal{K}_\varphi(f_k). \tag{3}$$

The Koopman operator is infinite-dimensional; however, it is linear with respect to the space of scalar mappings $f : \mathcal{M} \to \mathbb{R}$, unlike φ which is nonlinear. Therefore, we approximate the Koopman operator—and hence, the dynamics of the system—via a finite-dimensional linear operator (i.e., a matrix) [11]. Numerical approaches for forming and studying the spectra of such linear approximations to the Koopman operator include dynamic mode decomposition (DMD) [12]–[15] and autoencoders [16]–[18].

II. SET Data

The Social Embeddedness Tool (SET) is an extensive questionnaire aimed at assessing the embeddedness (i.e., how close an individual is to the center of the gang) and other related factors of gang-involved youth and young adults (ages 14–25).

978-1-7281-6251-5/20/$31.00 ©2020 IEEE

SET questions fall into four categories: Self, Family, Group, and Other Group. Questions in the Self category concern a individual's understanding of the gang in which they are involved, as well as personal characteristics such as behavioral norms and attitudes towards impulsive risk taking. The Family category focuses on a participant's connection to their family and its behaviors. The Group category includes questions that evaluate a participant's attachment to the gang and its nature, whereas Other Group contains such questions corresponding to one other social group in which a participant is intensively involved. The other group can be either prosocial or antisocial.

As of March 2019, 1,998 individuals completed a total of 3,416 SET questionnaires. Of those, 548 participants completed the questionnaire more than once, and 71 participants completed it three times or more. Due to the shortage of multiple SET-Retests for single participants, the present analysis focuses exclusively on discovering the dynamics in response to a first cycle of GRYD FCM services, meaning that only the SET-Intake and the first SET-Retest are used. A cycle of GRYD FCM services typically lasts six months. There is insufficient data to make reasonable conclusions about the effects of "Other Group" characteristics on individuals' gang involvement since only 57.66% of the 548 participants are members of another social group other than a primary gang and there is a great deal of diversity as to what these other groups are. Therefore, we discard the questions in this category and focus on the other three categories, which include 61 overall measurable questions. After removing responses with missing entries and errors, we have SET-Intake and SET-Retest questionnaires from 363 unique individuals. Of these, 62.53% are male, 37.47% are female, 34.71% 19 years of age or older, and 65.29% are below the age of 19.

For uniformity of the analysis, all the questions are rescaled so that 0 is always the lowest possible answer and that 1 is the highest possible answer. A lower answer represents less gang involvement and/or less gang activity. In this work, SET-Intake and SET-Retest data are arranged into matrices X and Y respectively. The i^{th} column vector of X and Y are the SET-Intake and SET-Retest from the i^{th} participant, and the j^{th} entry of the vector is the i^{th} participant's answer to the j^{th} question. Thus, we have X,$Y \in \mathbb{R}^{61\times 363}$.

III. DMD Analysis

First proposed by Schmid [12] to study fluid dynamics, dynamic mode decomposition (DMD) has gained popularity due to its semantic interpretability and ability to deal with non-linear dynamics. Although there are several state-of-the-art variants of DMD, these methods share a common goal of finding an appropriate matrix A such that $Y \approx AX$ given input data X and output data Y, i.e. A approximates the Koopman operator. Spectral analysis on A can then yield insights into the dynamics of the original system. In particular, the norm of an eigenvalue λ of A specifies the growth or decay of the corresponding eigenvector. In our case, $\|\lambda\|$ less than 1 suggests a decay in gang involvement, and $\|\lambda\|$ greater than 1 suggests an increase in gang involvement.

We note that a common alternative to DMD is principal component analysis (PCA), also called proper orthogonal decomposition (POD). DMD has been found to possess several advantages over PCA. For instance, [12] notes that DMD applies directly to data, while PCA only processes second-order statistics of a dataset. Moreover, [19] observes that DMD yields future state predictions for any time t, unlike PCA, which would require solves for each t. They also remark that DMD-based robust PCA can outperform L_1 optimization methods by 3-4 orders of magnitude [20].

A. Exact DMD

Exact DMD [13] provides an efficient way to solve the least-squares problem $Y = AX$, which has the analytical solution $A = YX^+$ (where X^+ is the right pseudoinverse of X). Instead of directly analyzing the eigendecomposition of A, which can be prohibitively computationally expensive, Exact DMD uses the singular value decomposition (SVD) to compute a reduced-order matrix $\tilde{A}$ whose eigenvalues belong to the spectrum of A and whose eigenvectors are connected to those of A by conjugate matrices. The algorithm is summarized in Algorithm 1.

Algorithm 1: Exact Dynamic Mode Decomposition

Input: $X, Y \in \mathbb{R}^{m\times n}$ and $r \in \mathbb{R}$

1 Compute the Singular Value Decomposition (SVD) of $X = U\Sigma V^*$.

2 We "trim" our basis to obtain approximations for input and output matrices. Define $\tilde{X} = U_r^* X$, $\tilde{Y} = U_r^* Y$ where U_r is the projection of U onto the first r modes.

3 Derive $\tilde{A} = \tilde{Y} V_r \Sigma_r^{-1}$ where V_r and Σ_r are projections of V and Σ to their first r modes respectively.

4 Compute the eigendecomposition of $\tilde{A}$.

We demonstrate the results of Exact DMD on the SET data for GRYD FCM participants in Figure 1, which plots the spectrum of A. Note that the real parts of the logarithm of all 30 eigenvalues are negative, suggesting a major decay of gang involvement and activity after a single cycle of GRYD FCM services.

B. Consistent DMD

Consistent DMD [21] is an alternative method to compute the dynamic mode decomposition evolution matrix A. The central difference of Consistent DMD is that it takes into account both forward and backward dynamics in order to diminish the effect of noise in the data. Mathematically, one can pose the dynamics of the system as

$$\text{Forward Dynamics: } Y = AX$$
$$\text{Backward Dynamics: } X = A^{-1}Y.$$

Then, Consistent DMD seeks an A which solves the optimization problem

$$\min_A \frac{1}{2}\left|AX - Y\right|_F + \frac{1}{2}\left|A^{-1}Y - X\right|_F,$$

where $|\cdot|_F$ is the Frobenius norm. By change of variables, the constrained minimization problem

$$\min_{A,B} \tfrac{1}{2}\left|AX - Y\right|_F + \tfrac{1}{2}\left|BY - X\right|_F \text{ such that } AB = I, BA = I$$

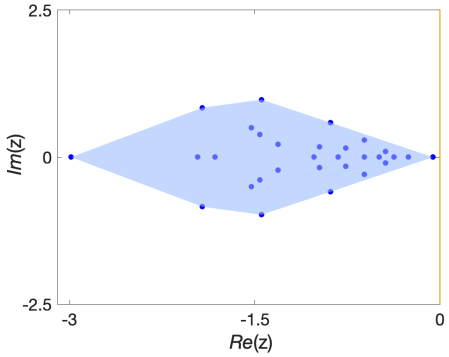

Fig. 1. Plot of the distribution of the logarithm of eigenvalues of A derived from Exact DMD. In this and similar plots throughout the paper, the shaded polygon indicates the convex hull of the points, which gives a general idea of the distribution of the eigenvalues and the dynamics. Positive real parts of the eigenvalue logarithms indicate growing modes, whereas negative real parts suggest decaying modes. The imaginary parts do not have a strong semantic interpretation on their own, but eigenvalues with larger overall norms contribute more to the dynamics of the system, as in similar techniques like principal component analysis.

can be solved efficiently using the Alternating Method of Multipliers [22]. This implementation of Consistent DMD is summarized in Algorithm 2.

Figure 2 shows that, when using Consistent DMD, all 30 eigenvalues of the evolution matrix A have a norm less than 1, which agrees with the results obtained via Exact DMD. However, there is a distinct difference between the range of $Re(log(\text{eigenvalue}))$ of Exact DMD and that of Consistent DMD. Nonetheless, the fact that all modes appear to be decaying with both methods is suggestive of the overall efficacy of GRYD FCM services.

Algorithm 2: Consistent Dynamic Mode Decomposition

Input: $X, Y \in \mathbb{R}^{m\times n}$ and $r, \rho \in \mathbb{R}$

Compute the SVD of $X = U\Sigma V^*$.

Define $\tilde{X} = U_r^* X$, $\tilde{Y} = U_r^* Y$ where U_r is the projection of U onto the first r modes.

Define $A^0 = \tilde{Y}\tilde{X}^+$, $B^0 = \tilde{X}\tilde{Y}^+$, $Q^0 = 0$, $Q \in \mathbb{R}^{2r\times r}$

for $k = 0, 1, 2, \ldots$ **do**

$A^{k+1} = \text{Sylvester}(C_1, C_2, C_3)$

$B^{k+1} = \text{Sylvester}(D_1, D_2, D_3)$

where $Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}$, $Q_1, Q_2 \in \mathbb{R}^{r\times r}$

$$\begin{cases} C_1 = \rho(B^k)^T B^k \\ C_2 = \tilde{X}\tilde{X}^T + \rho B^k (B^k)^T \\ C_3 = \tilde{Y}\tilde{X}^T + 2\rho(B^k)^T - \rho Q_1^k (B^k)^T - \rho(B^k)^T Q_2^k \\ D_1 = \rho(A^{k+1})^T A^{k+1} \\ D_2 = \tilde{Y}\tilde{Y}^T + \rho A^{k+1}(A^{k+1})^T \\ D_3 = \tilde{X}\tilde{Y}^T + 2\rho(A^{k+1})^T - \rho Q_1^k (A^{k+1})^T - \rho(A^{k+1})^T Q_2^k \end{cases}$$

$$Q^{k+1} = Q^k + \begin{pmatrix} A^{k+1}B^{k+1} - I \\ B^{k+1}A^{k+1} - I \end{pmatrix}$$

Update ρ for faster convergence.

end

Compute the eigendecomposition of A.

For a more detailed understanding of the effectiveness of the GRYD FCM Program, we study the separate dynamics

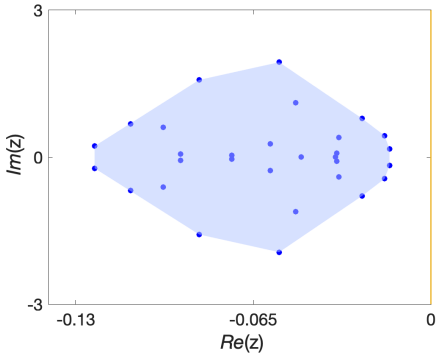

Fig. 2. Consistent DMD Log(Eigenvalue)

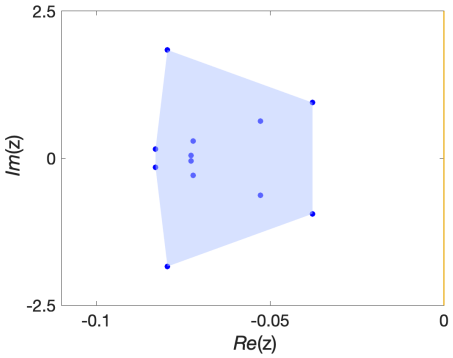

Fig. 3. "Self" Log(Eigenvalue)

of responses each of the three categories of questions: Self, Family, and Group.

1) Self: Figure 3 shows that the real part of the logarithm of every eigenvalue is negative, suggesting an overall decrease in the participants' risk level in the Self category. Hence, one can conclude that participants generally experienced positive self-growth between SET-Intake and SET-Retest evaluations.

2) Family: Figure 4 shows that though 7 out of 31 eigenvalues have norm greater than 1, the majority of the modes are clearly decaying; moreover, the eigenvalues with negative real components have larger magnitudes than those which have positive real components. Hence, we observe that participants generally experienced decreasing risk levels related to family life over the course of one GRYD FCM services cycle.

To better study the dynamics of the Family category, we conduct a subgroup analysis where participants are characterized based on their sex and age. Out of 363 participants, 153 are male aged 18 or younger, 74 are male aged 19 or older, 84 are female aged 18 or younger, and 52 are female aged 19 or older.

Figure 5 gives a visual representation of the dynamics of these subgroups. We observe that when the sex of a group is fixed, results for the younger group cluster around $Re(z) = 0$, whereas those for the elder group are distributed more widely across the plane. This suggests that older participants in GRYD FCM services have greater variance in outcomes, and also that

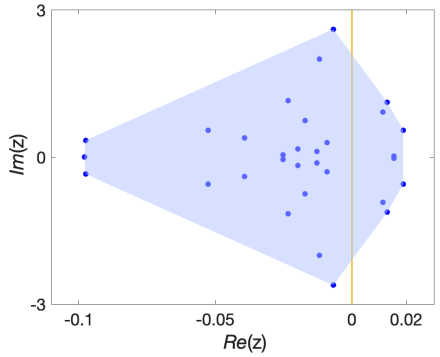

Fig. 4. "Family" Log(Eigenvalue)

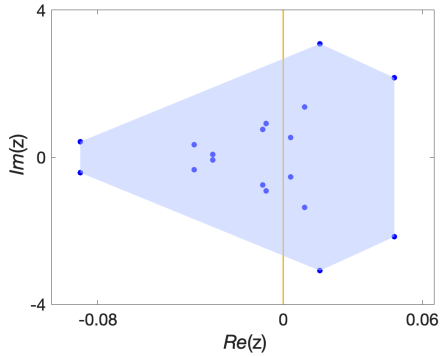

Fig. 6. "Group" Log(Eigenvalue)

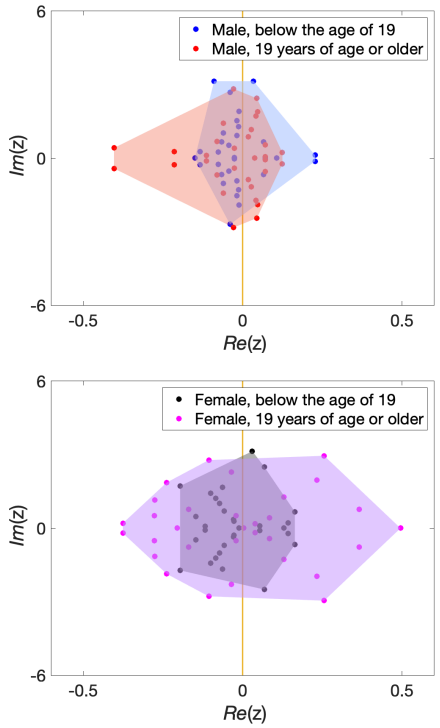

Fig. 5. Plot of the log of eigenvalues of the four groups of participants in the category of Family. The figure highlights the comparison between two groups that have the same sex but different age range.

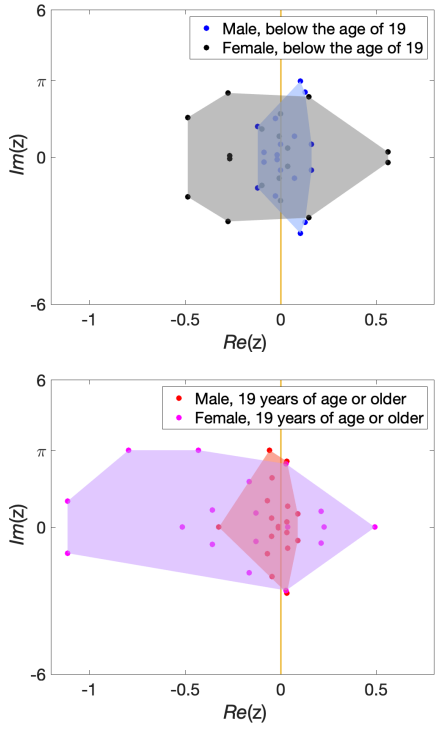

Fig. 7. Plot of the logarithm of eigenvalues of the four groups of participants in the category of Group. The plot visually demonstrates the comparison between two groups of the same age range but different sex. Note that while eigenvalues occur in complex conjugate pairs, logarithms of negative real eigenvalues will have a fixed imaginary part of πi and hence the shaded polygons may not be symmetric.

they can have more extreme changes (positive or negative) than what is observed for younger participants.

3) Group: The Group category displays more complex dynamics than the Self and Family categories. Figure 6 shows that 8 out of 18 eigenvalues have a norm greater than 1, and they do not cluster near $Re(z) = 0$, suggesting the presence of a non-negligible growing force. Hence we conclude that some features of risk associated with groups are growing, while others are decaying.

Figure 7 shows the subgroup dynamics of the category of Group. Compared with the Family category questions, in which age is significant in determining the dynamics, the primary distinction is by gender for the Group category questions. There is much greater variability in the dynamics for females in both age groups. Males are less likely to change in the Group category questions after the first cycle of GRYD FCM services in both age groups.

C. DMD with Control

DMD with Control [23] is a modified version of Exact DMD that incorporates the effect of control variables to disambiguate between the underlying dynamics and external actuation, e.g., GRYD FCM services. Instead of just dealing with the input matrix $X \in \mathbb{R}^{m \times n}$ and the output matrix $Y \in \mathbb{R}^{m \times n}$, a matrix

$\Upsilon \in \mathbb{R}^{l\times n}$ with snapshots of the state of the control variables introduced into the system. Mathematically, we aim to find the appropriate matrix A and B such that

$$Y \approx AX + B\Upsilon.$$

The obtained matrix A represents the dynamics when the intervention is hypothetically not present. This method is especially useful when it is impossible to get data free from the effect of the stimuli.

Algorithm 3: Dynamic Mode Decomposition with Control

Input: $X, Y \in \mathbb{R}^{m\times n}, \Upsilon \in \mathbb{R}^{l\times n}$ and $p, r \in \mathbb{R}$

Define $\Omega = \begin{pmatrix} X \\ \Upsilon \end{pmatrix}$

Compute the SVD of $\Omega = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \Sigma V^* \approx \begin{pmatrix} U_{1p} \\ U_{2p} \end{pmatrix} \Sigma_p V_p^*$

Compute the SVD of $Y = \tilde{U}\tilde{\Sigma}\tilde{V}^* \approx \tilde{U}_r\tilde{\Sigma}_r\tilde{V}_r^*$

Compute $A = Y\tilde{V}\tilde{\Sigma}^{-1}\tilde{U}_1^*$ and $B = Y\tilde{V}\tilde{\Sigma}^{-1}\tilde{U}_2^*$

Compute the eigendecomposition of A

For the GRYD dataset, we assume there are underlying dynamics for individuals' responses when intervention is absent and that GRYD FCM services are the actuation. To ensure the changes in individuals' responses, before and after a program cycle, are not solely due to the effect of time, we compare the dynamics derived from Exact DMD and that from DMD with Control. In this work, we use each of the five categories of GRYD treatment strategies as control variables. These five categories are mentorship for school/job readiness, substance abuse/anger management/life skill classes, community service, "40 Developmental Assets," and counseling service. For every category, if the participant has adopted the strategy, 1 is assigned to the corresponding entry in Υ, and if not, 0 is assigned. Mathematically we have

$$\mathbb{R}^{5\times n} \ni \Upsilon = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix} \text{ where } v_i = \begin{bmatrix} \text{Mentorship} \\ \text{Life Skill} \\ \text{Community} \\ \text{40 DA} \\ \text{Counseling} \end{bmatrix}$$

For example, if $v_i = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \end{bmatrix}^T$, then the i^{th} participant has received mentorship for school and work and has participated in community service. We note that not every participant has a record of the strategies they took during the first cycle of the program; out of the 363 participants who completed the SET-Intake and SET-Retest, only 204 of them have data for Υ.

Figure 8 shows that all the eigenvalues derived from Exact DMD and DMD with Control have norm less than 1, suggesting overall decaying dynamics for both actual and hypothetical scenarios. It is also evident that the logarithm of eigenvalues derived from DMD with Control distribute more closely to $Re(z) = 0$, which means that GRYD FCM services contribute to a greater reduction of gang involvement and activity after controlling for treatment strategies. One

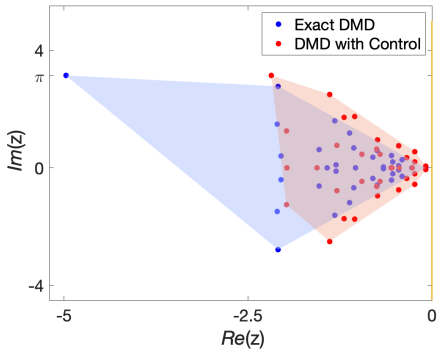

Fig. 8. Plot of the distribution of logarithm of eigenvalues of A derived from Exact DMD and DMD with Control. The blue markers and light blue region visually represent the dynamics when GRYD is viewed as endogenous for the system. The red markers and red region show the dynamics of the same participants when GRYD is hypothetically not in action.

may conclude that the GRYD FCM Program is succeeding in transferring attachments from gangs to positive activities.

1) Statistical Analysis: To quantitatively study the effectiveness of GRYD FCM services, we conduct a statistical analysis on the obtained dynamics. For every SET-Intake response x_i, the predicted SET-Retest response when GRYD FCM services are hypothetically not in action is calculated and denoted as $\bar{y}_i$ with $\bar{y}_i = \bar{U}\bar{A}\bar{U}^* x_i$, where $\bar{A}$ and $\bar{U}$ represent matrices A and $\tilde{U}_r$ derived from DMD with Control. Thus, we have two samples of SET-Retest responses, the first being the actual SET-Retest answers from participants, and the second being the set of $\bar{y}_i$. We then conduct a kernel two-sample test [24] on the data. The kernel two-sample test uses maximum mean discrepancy (MMD) as the test statistic, and when MMD is greater than the threshold derived from bootstrapping, the null hypothesis is rejected. In our case, we have a statistic of 8.6259 and a threshold of 2.1259, suggesting that the difference between the actual SET-Retest responses and the hypothetical ones is statistically significant. Thus, according to this analysis, GRYD FCM services are statistically significantly effective in reducing gang involvement and activity.

IV. Autoencoder Analysis

The DMD algorithms discussed so far enable semantic interpretation of the modes of the approximate Koopman operator. However, a related task is also of interest: predicting individual's SET-Retest scores from their SET-Intake responses. This yields another view of the effectiveness of GRYD FCM services. For this task, we consider an additional approach for finding a finite dimensional approximation to the Koopman operator, namely, autoencoders. We assume there exists a (non-linear) transformation $\chi : \mathcal{M} \to \mathbb{R}^\kappa$ that maps the data from the observation space into a κ-dimensional latent space in which the Koopman operator can be approximated. Thus, we hope to find such a transformation χ and a real matrix $C \in \mathbb{R}^{\kappa\times\kappa}$ such that

$$\mathcal{K}_\varphi \approx \chi^{-1} \circ C \circ \chi, \tag{4}$$

or equivalently,

$$C \approx \chi \circ \mathcal{K}_\varphi \circ \chi^{-1}. \tag{5}$$

Therefore, our autoencoder network must both (1) accurately encode all observations in a κ-dimensional latent space and (2) discover a matrix that accurately encodes the dynamics of the system in the latent space.

In practice, such a transformation χ and its inverse χ^{-1} can be found via deep learning, by defining neural networks χ_e and χ_d (the encoder and decoder, respectively) such that $\chi_e \approx \chi$ and $\chi_d \approx \chi^{-1}$. Then, given a set of observations of the system f_k at time k, we can predict future observations one step forward in time by

$$\hat{f}_{k+1} = \chi_d \circ C \circ \chi_e(f_k). \tag{6}$$

A. Consistent Dynamics

Similar to DMD, to add consistency to our Koopman autoencoder, we assume there exist backward dynamics that can be encoded in the latent space by a matrix $D \in \mathbb{R}^{\kappa\times\kappa}$, where $D \approx C^{-1}$. Thus, we can generate backward predictions by

$$\check{f}_{k-1} = \chi_d \circ D \circ \chi_e(f_k). \tag{7}$$

In Section IV-F, we investigate whether introducing these consistent dynamics increases the predictive power of our model.

B. Loss Function

The model parameters (the weights and biases of χ_e and χ_d, and the entries of C and D) were trained according to a loss term consisting of four terms. The first of these is the identity loss

$$\mathcal{E}_{\text{id}} = \frac{1}{2n}\sum_{k=1}^{n} \|f_k - \tilde{f}_k\|_2^2, \tag{8}$$

where n is the total number of observations and $\tilde{f}_k$ is the k^{th} observation as reconstructed by the encoder and decoder network, i.e.

$$\tilde{f}_k = \chi_d \circ \chi_e(f_k). \tag{9}$$

This helps maintain the constraint that $\chi_d \circ \chi_e \approx \text{id}$, so that the autoencoder accurately encodes the data in the latent space.

Next, there are forward and backward loss terms, which measure the error of the model's forward and backward predictions on the training data. Specifically, these are

$$\mathcal{E}_{\text{fwd}} = \frac{1}{2n}\sum_{k=1}^{n} \|\hat{f}_{k+1} - f_{k+1}\|_2^2 \tag{10}$$

and

$$\mathcal{E}_{\text{bwd}} = \frac{1}{2n}\sum_{k=1}^{n} \|\check{f}_{k-1} - f_{k-1}\|_2^2. \tag{11}$$

Finally, the consistency loss term, which ensures that the forward and backward dynamics are consistent (i.e., that $CD \approx DC \approx I_\kappa$), is given by

$$\mathcal{E}_{\text{con}} = \sum_{k=1}^{\kappa} \frac{1}{2k}\|D_{k*}C_{*k} - I_k\|_F^2 + \frac{1}{2k}\|C_{k*}D_{*k} - I_k\|_F^2, \tag{12}$$

where the subscripts $_{k*}$ and $_{*k}$ denote the first k rows and the first k columns of a matrix, respectively, and $\|\cdot\|_F$ is the Frobenius norm.

The total loss is then given by

$$\mathcal{E} = \lambda_{\text{id}}\mathcal{E}_{\text{id}} + \lambda_{\text{fwd}}\mathcal{E}_{\text{fwd}} + \lambda_{\text{bwd}}\mathcal{E}_{\text{bwd}} + \lambda_{\text{con}}\mathcal{E}_{\text{con}}. \tag{13}$$

For our model, we chose to use the weights $\lambda_{\text{id}} = \lambda_{\text{fwd}} = 1$, $\lambda_{\text{bwd}} = 0.1$, and $\lambda_{\text{con}} = 0.01$, as our primary focus is placed on the validity of the autoencoder and the accuracy of the forward predictions.

To determine whether introducing consistent dynamics improves the performance of the model, we also train a model where we set $\lambda_{\text{bwd}} = \lambda_{\text{con}} = 0$, effectively neglecting the role of backward dynamics.

C. Network Architecture

The encoder and decoder layers of our model consist of three fully-connected layers with $\tanh$ activation. These three hidden layers map the data from dimension m to p, p to p, and p to κ respectively for the encoder and vice versa for the decoder, where m is the dimension of the input (the number of questions per individual), p is a hidden layer dimension, and κ is the dimension of the latent space.

To determine the optimal dimension of the latent space κ, we use repeated cross-validation. We train the model on 10 random 80%-20% train-test splits according to Section IV-E, varying κ between 2 and 29. To avoid an exhaustive search on (κ, p), and so that the total number of parameters in the encoder and decoder networks would roughly scale with κ, we fix $p = 2\kappa$. As shown in Figure 9, the model's mean squared error (MSE) on validation data over the 10 repeated trials stops improving significantly after $\kappa = 15$, so this is the latent space dimension we choose for our model.

Likewise, to select the hidden layer size p, we plot the model's mean validation loss across 10 trials as we vary p, with the latent space dimension fixed at $\kappa = 12$. We see in Figure 10 below that $p = 24$ is a reasonable choice for the size of the hidden layers in the autoencoder.

D. Autoencoder with Control

Similar to DMD, we implement controls in our autoencoder framework by introducing certain demographic variables into the model. The additional information may allow the model to better determine the dynamics for the system as a whole, as it is likely that GRYD FCM services have varying effects on individuals of different backgrounds. Specifically, we incorporate each participant's age and sex as additional inputs to the model, using one-hot encoding for age (sex was binary, and thus left as a single column). In total, there are 18

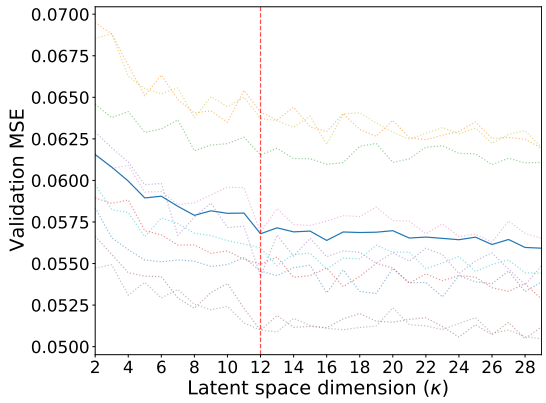

Fig. 9. Mean validation MSE (solid blue line) vs. latent space dimension (κ) across 10 repeated validation runs (pale dotted lines). Improvements are insignificant past $\kappa = 12$ (vertical dashed line).

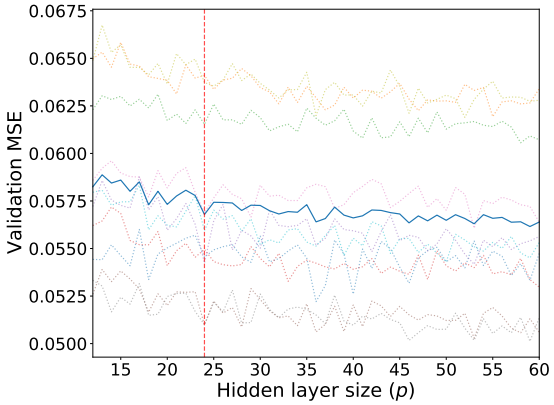

Fig. 10. Mean validation MSE (solid blue line) vs. hidden layer size (p) across 10 repeated validation runs (pale dotted lines). Improvements are insignificant past $p = 24$ (vertical dashed line).

unique ages encoded as binary control variables. Aside from these additional inputs, the rest of the model, including the architecture and training method, is unchanged.

E. Training Method

We initialize the model such that C and D are random orthogonal matrices generated from the $O(\kappa)$ Haar distribution, which acts as a uniform distribution on $\mathbb{R}^{\kappa\times\kappa}$ [25]. This promotes the stability of the model [11], as the eigenvalues of C and D initially lie on the unit circle in $\mathbb{C}$. The model is trained using the AdamW optimizer [26] with gradient clipping and L2 regularization. We use an initial learning rate of 0.01, set to decay every 50 epochs. For each trial, we randomly split the data (the SET-Intake SET-Retest response pairs) into 80% training and 20% testing sets, and then minimize $\mathcal{E}$ over 600 epochs. At every epoch, we also calculate the validation loss as the mean squared error of forward predictions on the testing set, akin to $\mathcal{E}_{\text{fwd}}$ in Eq. (10). At the end of the training epochs, we revert the model to the epoch that minimized validation loss, effectively implementing after-the-fact early stopping to avoid overfitting.

F. Results

To assess the predictive power of the various states of our model, we train four different configurations, representing combinations of control/no control and consistent dynamics/no consistent dynamics. We train the model over 100 trials, with each trial representing a different 80%-20% train-test data split, and record the minimum mean squared error (MSE) of the various model configurations on the testing data. The mean and standard deviation of the test MSE across 100 trials for each model configuration are recorded in Table I.

TABLE I
MEAN AND STANDARD DEVIATION OF VALIDATION MSE (OVER N=100 TRIALS), BY AUTOENCODER CONFIGURATION

		Consistent dynamics	
		Yes	*No*
Control	*Yes*	Mean: 0.05028 SD: 0.00416	Mean: 0.05050 SD: 0.00404
	No	Mean: 0.04926 SD: 0.00412	Mean: 0.04924 SD: 0.00426

The results in Table I show that the model configuration with the lowest validation MSE was the autoencoder with no consistent dynamics and no control. However, to test whether this result is statistically significant, we perform a two-way repeated measures ANOVA. Each of the 100 trials is treated as a subject (as it is likely that the exact 80%-20% data split has a non-negligible effect on the performance of each model), consistent dynamics and control are treated as within-subjects factors (with two levels each, yes and no), and validation MSE is treated as the dependent variable. ANOVA is based on two assumptions: (1) that the samples are drawn from Normal distributions, and (2) that the variances of the differences between all possible combinations of within-subject factors are equal, i.e. sphericity. [27] raises concern over ANOVA's use in cases where these assumptions are violated. However, our data passes the Shapiro-Wilk test of normality for all possible combinations of the within-subject factors ($p > 0.05$) as shown in Table II, indicating that we cannot reject the null hypothesis that the validation MSEs are drawn from normal distributions for each possible model configuration. Second, our experiment automatically passes Mauchly's sphericity test, as sphericity is always met for within-subjects factors with two levels [28].

TABLE II
SHAPIRO-WILK TEST OF NORMALITY

	Consistent	Control	variable	statistic	p
1	0	0	MSE	0.984	0.287
2	0	1	MSE	0.985	0.309
3	1	0	MSE	0.983	0.221
4	1	1	MSE	0.987	0.449

Therefore, we report the two-way repeated-measures AVNOA test results for the within-subject effects in Table III. Evidently, of the three effects, only Control is statistically significant ($p < 0.05$), so we may only reject the null hypothesis that the model has equal predictive power with and without control, and cannot reject the null hypotheses

TABLE III
TWO-WAY REPEATED MEASURES ANOVA

Effect	df	F value	p value
Consistent	1	1.70	0.193
Control	1	212	1.31×10^{-36}
Consistent:Control	1	2.46	0.118

that the model has equal predictive power with and without consistent dynamics or the interaction of consistent dynamics with control. Referring back to Table I, we see that adding control decreases predictive power (increases validation MSE). Thus, for the sake of minimizing the number of parameters required for training, we can consider the configuration with no consistent dynamics and no control to be the best of the four possible configurations in terms of predictive power.

G. Prediction using DMD and Autoencoder Methods

TABLE IV
MEAN AND STANDARD DEVIATION OF MSE FOR TEST DATA AND ALL DATA OVER 100 TRIALS

	Exact DMD	Consistent DMD	Autoencoder
MSE (Test)			
Mean	0.0453	0.0557	0.0492
Standard Deviation	0.0041	0.0056	0.0043
MSE (All)			
Mean	0.0387	0.0472	0.0444
Standard Deviation	1.7392×10^{-4}	2.6328×10^{-4}	0.0028

As can be seen in Table IV, for the task of predicting SET-Retest values from SET-Intake questionnaire responses, autoencoder methods outperform Consistent DMD on test data over 100 trials. Averaged over all data (testing and training), the autoencoder also outperforms Consistent DMD in terms of MSE, though in all cases exact DMD slightly outperforms autoencoders. Although the autoencoder model has strong predictive power (which could possibly be increased with additional training, deeper networks, etc.), the DMD models have the advantage of simplicity and interpretability. For example, the spectral analysis in the DMD model gives insight into the dynamics. With autoencoders, any matrix eigenvalues only give information about dynamics in the latent space, which may be difficult to translate back to actual dynamics. Therefore, one might prefer different methods for different tasks; e.g., using autoencoders to predict SET-Retest scores, but using DMD techniques for understanding the dynamics of the modes of the system.

V. DISCUSSION AND FUTURE WORK

This paper studies the dynamics of GRYD FCM services using Koopman operator theory. Growing and decaying modes of the system are analyzed using several variants of dynamic mode decomposition, and autoencoders are used to predict participants' retest scores on the SET questionnaire from their intake responses. Applying these numerical methods yields several statistically significant conclusions that suggest the efficacy of the GRYD FCM Program.

There are several opportunities for improving upon the present work. For example, we conduct our analysis using only complete samples in the SET data, which restricts our effective population size and predictive power. Data imputation techniques [29] like mean imputation or the recently-proposed GAIN method [30] could be considered to utilize incomplete samples (while avoiding biasing the imputed data). Further analysis could consider noise and biases in DMD-based models, for example as in [31], [32]. Moreover, the techniques presented in this paper could be applied to related datasets and issues, e.g. sociological processes which can be modeled as dynamical systems. The literature describes a number of datasets related to gangs and youth delinquency, e.g. [33]–[35], that may be good candidates for analysis using Koopman theory. Future applications may also include reducing biases in pretrial risk assessment [36], [37] or evaluating the effectiveness of various delinquency deterrence programs [38], [39].

ACKNOWLEDGMENTS

Omri Azencot provided preliminary code for Consistent DMD.

The SET was created by the City of Los Angeles Mayor's Office of Gang Reduction and Youth Development and is the copyright of the City of Los Angeles and has been used with permission. Any opinions, findings, conclusions or recommendations expressed in this study, however, are those of the authors and do not necessarily reflect the views of the GRYD Office. This research was funded by the City of Los Angeles contract number C-132202 ("GRYD Research and Evaluation").

REFERENCES

[1] Q. C. Thurman, A. L. Giacomazzi, M. D. Reisig, and D. G. Mueller, "Community-based gang prevention and intervention: An evaluation of the neutral zone," *Crime & Delinquency*, vol. 42, no. 2, pp. 279–295, 1996.

[2] J. Winterdyk and R. Ruddell, "Managing prison gangs: Results from a survey of U.S. prison systems," *Journal of Criminal Justice*, vol. 38, no. 4, pp. 730–736, 2010.

[3] M. Cahill and D. Hayeslip, "Findings from the evaluation of OJJDP's Gang Reduction Program," *Office of Juvenile Justice and Delinquency Prevention Juvenile Justice Bulletin*, 2010.

[4] I. A. Spergel, *Reducing youth gang violence: The little village gang project in Chicago*. Rowman Altamira, 2007.

[5] G. D. Gottfredson and D. C. Gottfredson, "Gang problems and gang programs in a national sample of schools," 2001.

[6] A. Tremblay, D. Herz, R. Zachery, and M. Kraus, "The Los Angeles Mayor's Office of Gang Reduction and Youth Development comprehensive strategy," *GRYD Research Brief No. 1*, 2020.

[7] B. O. Koopman and J. v. Neumann, "Dynamical systems of continuous spectra," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 18, no. 3, p. 255, 1932.

[8] B. O. Koopman, "Hamiltonian systems and transformation in Hilbert space," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 17, no. 5, p. 315, 1931.

[9] M. Budišić, R. Mohr, and I. Mezić, "Applied koopmanism," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 4, p. 047510, 2012.

[10] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, "Spectral analysis of nonlinear flows," *Journal of fluid mechanics*, vol. 641, no. 1, pp. 115–127, 2009.

[11] O. Azencot, B. Erichson, V. Lin, and M. W. Mahoney, "Forecasting sequential data using consistent Koopman autoencoders," in *Proceedings of the 37th International Conference on Machine Learning*. Vienna, Austria: PMLR, 2020.

[12] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," *Journal of Fluid Mechanics*, vol. 656, pp. 5–28, 2010.

[13] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, "On dynamic mode decomposition: Theory and applications," *Journal of Computational Dynamics*, vol. 1, no. 2, pp. 391–421, 2014.

[14] P. J. Schmid, L. Li, M. P. Juniper, and O. Pust, "Applications of the dynamic mode decomposition," *Theoretical and Computational Fluid Dynamics*, vol. 25, no. 1, pp. 249–259, 2011.

[15] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.

[16] N. Takeishi, Y. Kawahara, and T. Yairi, "Learning Koopman invariant subspaces for dynamic mode decomposition," in *Advances in Neural Information Processing Systems*, 2017, pp. 1130–1140.

[17] B. Lusch, J. N. Kutz, and S. L. Brunton, "Deep learning for universal linear embeddings of nonlinear dynamics," *Nature communications*, vol. 9, no. 1, pp. 1–10, 2018.

[18] S. E. Otto and C. W. Rowley, "Linearly recurrent autoencoder networks for learning dynamics," *SIAM Journal on Applied Dynamical Systems*, vol. 18, no. 1, pp. 558–593, 2019.

[19] J. N. Kutz, X. Fu, and S. L. Brunton, "Multiresolution dynamic mode decomposition," *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 2, pp. 713–735, 2016.

[20] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 1–37, 2011.

[21] O. Azencot, W. Yin, and A. L. Bertozzi, "Consistent dynamic mode decomposition," *SIAM Journal on Applied Dynamical Systems*, vol. 18, no. 3, pp. 1565–1585, 2019.

[22] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *Journal of Scientific Computing*, 2019.

[23] J. L. Proctor, S. L. Brunton, and J. N. Kutz, "Dynamic mode decomposition with control," *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 1, pp. 142–161, 2016.

[24] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Scholkopf, and A. Smola, "A kernel two-sample test," *Journal of Machine Learning Research*, vol. 13, no. 1, pp. 723–773, 2012.

[25] F. Mezzadri, "How to generate random matrices from the classical compact groups," 2007. [Online]. Available: https://arxiv.org/abs/math-ph/0609050v2

[26] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *ICLR 2019*. New Orleans, United States: IMLS, 2019.

[27] J. Demsar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, no. 1, pp. 1–30, 2006.

[28] P. R. Hinton, I. McMurray, and C. Brownlow, *SPSS Explained*. London, UK: Taylor & Francis, 2004.

[29] B. Efron, "Missing data, imputation, and the bootstrap," *Journal of the American Statistical Association*, vol. 89, no. 426, pp. 463–475, 1994.

[30] J. Yoon, J. Jordon, and M. Schaar, "GAIN: Missing data imputation using generative adversarial nets," in *International Conference on Machine Learning*, 2018, pp. 5689–5698.

[31] D. Duke, J. Soria, and D. Honnery, "An error analysis of the dynamic mode decomposition," *Experiments in fluids*, vol. 52, no. 2, pp. 529–542, 2012.

[32] M. S. Hemati, C. W. Rowley, E. A. Deem, and L. N. Cattafesta, "De-biasing the dynamic mode decomposition for applied Koopman spectral analysis of noisy datasets," *Theoretical and Computational Fluid Dynamics*, vol. 31, no. 4, pp. 349–368, 2017.

[33] D. J. Williams, D. Currie, W. Linden, and P. D. Donnelly, "Addressing gang-related violence in Glasgow: a preliminary pragmatic quasi-experimental evaluation of the community initiative to reduce violence (cirv)," *Aggression and violent behavior*, vol. 19, no. 6, pp. 686–691, 2014.

[34] P. Boxer, J. Kubik, M. Ostermann, and B. Veysey, "Gang involvement moderates the effectiveness of evidence-based intervention for justice-involved youth," *Children and youth services review*, vol. 52, pp. 26–33, 2015.

[35] M. J. Mayer, "Structural analysis of 1995–2005 School Crime Supplement datasets: factors influencing students' fear, anxiety, and avoidant behaviors," *Journal of School Violence*, vol. 9, no. 1, pp. 37–55, 2009.

[36] A. Milgram, A. M. Holsinger, M. Vannostrand, and M. W. Alsdorf, "Pretrial risk assessment: Improving public safety and fairness in pretrial decision making," *Federal Sentencing Reporter*, vol. 27, p. 216, 2014.

[37] S. L. Desmarais, S. A. Zottola, S. E. Duhart Clarke, and E. M. Lowder, "Predictive validity of pretrial risk assessments: A systematic review of the literature," *Criminal Justice and Behavior*, 2020.

[38] A. A. Braga, D. Weisburd, and B. Turchan, "Focused deterrence strategies and crime control: An updated systematic review and meta-analysis of the empirical evidence," *Criminology & Public Policy*, vol. 17, no. 1, pp. 205–250, 2018.

[39] N. Schell-Busey, S. S. Simpson, M. Rorie, and M. Alper, "What works? A systematic review of corporate crime deterrence," *Criminology & Public Policy*, vol. 15, no. 2, pp. 387–416, 2016.