For office use only	Team Control Number	For office use only
T1 _____	**1901997**	F1 _____
T2 _____		F2 _____
T3 _____	Problem Chosen	F3 _____
T4 _____	**C**	F4 _____

2019
MCM/ICM
Summary Sheet

Summary

Opioid abuse and addiction cause enormous harm to individuals and communities. Currently, the United States is experiencing a national crisis regarding with opioid abuse and addiction, which calls for new opioid policies to control the spread of opioid abuse. In this paper, we propose a mathematical model to simulate the procedure of drug transmission and discover the correlation between drug transmission and socio-economics, as well as make recommendations to the Chief Administrator in DEA/NFLIS.

We first construct a network of counties based on their adjacency relationships. In this network, a node denotes a county, and an edge is established if two counties are adjacency. Then, we propose a new graph propagation model to model the procedure of drug transmission. Unlike the traditional infectious disease models, our model considers the topological relationships, distances, and influence among counties. Specifically, the number of drug identifications in a county is determined by two factors, which are 1) the number of drug identifications in the last year in the same county (internal factor), and 2) the number of drug identifications of its adjacency counties (external factor). We develop a linear regression model to model the influence of adjacency counties.

We conduct extensive experiments on the NFLIS dataset to evaluate the model. The results indicate that our model is very effective to model the ground-truth trend of opioid transmission. By a sensitive analysis for our model, we find that the number of drug cases is mainly determined by the internal factor. That is to say, drugs mainly spread within the county, and the spread between two counties is not very significant.

Second, we devise a new CountyRank algorithm, based on the well-known PageRank algorithm, to identify the source of drug transmission on the network. On the basis of the CountyRank algorithm, we are able to compute a rank for each county, in which the high-rank county could be of a source of drug transmission. The results on NFLTS dataset indicate that the source of PA state may be at PHILADELPHIA, ALLEGHENY;...

Third, we make use of our model to predict the spread of drugs in various states. Based on our results, in 2020, the US government needs to pay particular attention to the drug crisis in PHILADELPHIA state. In addition, we also introduce socio-economic factors of counties to improve the proposed model. By a careful correlation analysis, we obtains seven socio-economic features from 600 features for modeling, which are the most relevant to the drug cases. The results show that marriage and ethnicity have a more significant impact on drug transmission.

Finally, we make the following recommendations for the states to control the spread of drugs: increase the science of drugs, strengthen the psychological counseling of the people, raise the concern for social stability, and strengthen supervision to make people equal. Experimental results confirm the validity of these recommendations.

Keywords: **Graph Propagation Model, Linear Regression, Drug Spread, PageRank**

MEMO

To: the Chief Administrator, DEA/NFLIS
From: Team# 1901997
Date: 29 January 2018
Subject: On opioid crisis

Purpose

We have conducted a detailed analysis of the reported data, and would like to summarize the results on the spread of drug from 2010 to 2017 and predicting the trend of drug transmission in the absence of any policy changes, as well as giving our recommendations.

Summary for the Drug Spread during 2010-2017

Based on the number of drug cases reported between 2010 and 2017 and the simulation results of our model, we find the following:

- The spread of drugs is growing, often propagating to the periphery centering in a county.
- The number of drugs in the next year of a county relies mainly on internal factors, namely, the number of drug cases, the marital status of residents, and the distribution of ethnic groups. At the same time, the drug transmission in a county in the next year is also affected by drug cases in its neighboring counties.
- Among the effects of internal factors, the influences of the Marriage and race factor is the most significant factors.

Prediction the Trend of Drug Transmission

Assume that the current drug policies keep unchanged. Based on this assumption, we propose a model to predict the most widespread and severely affected counties in five states. The following conclusions were obtained:

- By 2020 year, the first county, namely the PHILADELPHIA County, in which the heroin is reported in the PA state is greater than the threshold. This is a serious beginning of the situation.
- By 2035 year, all states have heroin reported cases with a number of cases greater than the threshold. The situation of drug spread broke out in that year.

Recommended Goals and Actions

According to the factors above and the characteristics of drug spread, we offer the following suggestions to overcome the problems:

- It is important to make an effort to let the public to know the harm of opium and other addictive drugs to human beings. Pursuing a short and meaningless pleasure requires a considerable price.

- Strengthen the psychological counseling of people; try to give emotional care; improve the quality of community psychologists; and provide more professional guidance and advice to those in need.
- Raise attention to the degree of social stability. When the number of unemployed in the society is high or is increasing rapidly, relevant government departments should strengthen the supervision of addictive drugs and alleviate the number of unemployed.
- Equality of ethnicity and increased oversight in places where particular ethnic groups gather.

Contents

1 Introduction

1.1 Background

Drug abuse is now rampant in the United States. The use of drugs has gone beyond traditional medicine for analgesics, and more people are using it to get special pleasure for recreational purposes. What's more, the phenomenon of drug abuse itself has the nature of contagion. Drug abuse inevitably leads to severe health and social problem and increases financial expenditure in the United States. Specifically, the mistreatment of the drug can bring various diseases to the public and aggravate the spread of diseases. It may also breed drug crimes, such as robbery, violence, theft and so on, which has a negative impact on public security and public safety. [8]

Now, the growing problem of drug abuse has attracted the attention of the United States government, and many organizations are trying to solve the problem. The key point to solve the problem of drug abuse is to analyze spread of drugs and the influencing factors. Only in this way can the government put forward effective strategies and put an end to drug abuse.

1.2 Restatement of the Problem

For address the problem of drug abuse, we need to consider three aspects, namely, drug transmission mode, the influencing factors, and effective strategies. The questions we are required to answer are as follows.

- Based on the information network of drug cases, establish a model to describe the condition and characteristics of drug transmission. Furthermore, the model can determine the likely source of drug spread and predict the future drug use.
- Analyze the causes of the drug's transmission status, and identify the factors that have a more significant correlation with drug spread according to the economic data. Then, study the impact of those factors and optimize the propagation model.
- Combined with the above model and analysis results about drug transmission, give a solution to the drug abuse status and verify the success of the strategy.
- Write a 1-2 page memo to the Chief Executive to determine the key factors they should consider.

1.3 Our Work

As described above, the key to solving the problem of drug abuse lies in analyzing its spread mode and influencing factors, and then proposing effective strategies to deal with the drug abuse crisis. In this paper, we first build a basic spread model and then optimize it by adding important factors. Finally, we give a solution to the drug abuse status and verify the success of the strategy. The main work of this paper is as follows.

- Basic Spread Model. A model is developed to describe the spread and features of the reported synthetic opioid and heroin incidents. [3] We innovatively improve

the PageRank algorithm to find the source of drug transmission. In addition, We also predict the future trend of drug spread.

- The Refined Transmission Model. We study the economic data and identify some factors that are highly relevant to drug transmission. And then the improved graph propagation model is introduced by considering these socio-economic factors.
- Strategy. Some strategies are designed to address the drug abuse crisis. And, we also apply models to assess the importance of these strategies.

The rest of this paper is organized as follows. We make some strong assumptions to simplify the graph propagation model and present the notations in Section 2. In Section 3, Section 4 and Section 5 are used to introduce the three main works mentioned above. We also give the strengths and weaknesses of the model in Section 6. Finally, we conclude the paper in Section 7.

2 Model Assumptions and Notation

In this section, we introduce some assumptions to simplify the course of modeling and draw some reasonable conclusions. Meanwhile, for ease of description, we also present the notations used in this paper.

2.1 Assumptions

In our model, we make the following assumptions.

- **Government policies remain unchanged.** In the past few years and in the future, the intensity of restricting drug abuse is constant. Namely, the policy of each state will not change. This assumption is essential for predicting.
- **The situation in each county in a state is the same.** In the same state, all counties have the same attitude toward drug abuse, and citizens have the same mobility.
- **Citizens in neighboring counties are free to move.** There is no restriction on the movement of people or drug users.
- **The number of reported drug-related cases is proportional to the number of drug users.** When the amount of drug-related incidents in a county increases, it means that the number of drug users in this county is rising.
- **The farther the distance between the counties, the smaller the impact on the other party** The farther the county A from county B, the less likely the citizens of the county A are to interact with the citizens of B. The more difficult it is for the county A to influence the drug cases in the county B.

2.2 Notation

Here we list the symbols and notations used in this paper, as illustrated in Table 1.

Table 1: Symbols and Notations

Notation	Meaning
u	predicting county
v_i	i-level neighbor of u
t	year
$\mathrm{g}_t(u)$	In u, the amount of cases in t year
p_i	a node in the network
D_i	a collection of all nodes linked into node p_i
$L(p_i)$	Out-degree of node p_i
q	Probability of entering the node from neighboring nodes
$1-q$	Probability of random walk
N	Total number of nodes
$sum(p_i)$	Number of drug cases on node p_i
α	Intrinsic factor
β	First-order external factor
M	First order adjacency matrix
I_N	N-order unit equation
T	Probability transfer matrix
$CountyRank0$	The initial value of CountyRank
CR	the CountyRank value
λ_i	Internal parameter
e_i	An internal factor
$f(e_i)$	value of the county's e_i

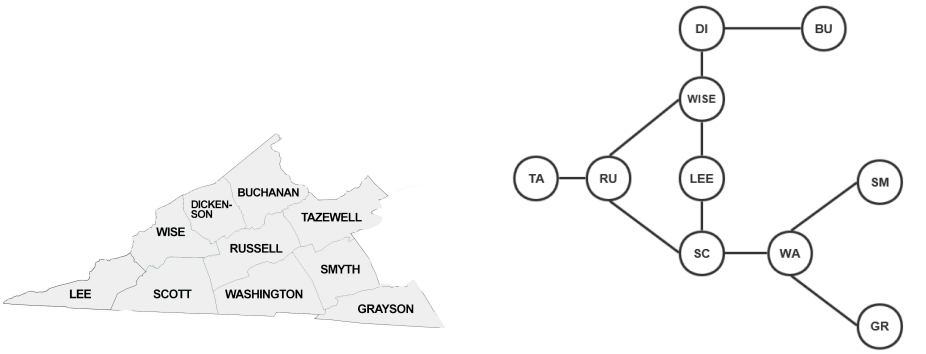
3 The Basic Spread Model

In this section, we first develop the Basic Spread Model (BSM) to describe the transmission of drug in and between states and countries. Compared with the widely used infectious disease models, [1] we consider the distances and relationships between counties to simulate the circulation and influence between counties. Next, linear regression is employed to fit and find the appropriate parameters of the BSM model. [11]By using the BSM model and the parameters obtained by linear regression, we infer that the spread of drugs has the characteristics of "the rich get richer". Then, we design an improved PageRank algorithm, named as CountyRank, to trace the source of drug abuse. Finally, we predict the development of drug abuse, including when and where drug will cause concern to the goverment in the future.

3.1 Graph Propagation model

Before introducing our BSM model, we give the organization method of network which is the foundation of constructing the BSM model. [2]The specific methods are as follows. We model a network for a state and each county in this state acts as a node in the network. If one county borders another county geographically, we consider that an edge connects the two nodes. In this way, we can transform a state into a network. [5] Figure 1(a) illustrates the geographical location of the LEE county of VA State. And Figure 1(b) is the model network of counties near the LEE county of VA State by the above method. [11] Based on the network, we abstract the process, which drugs spread between states and counties, as the process by which data rises in nodes and flows between nodes and nodes

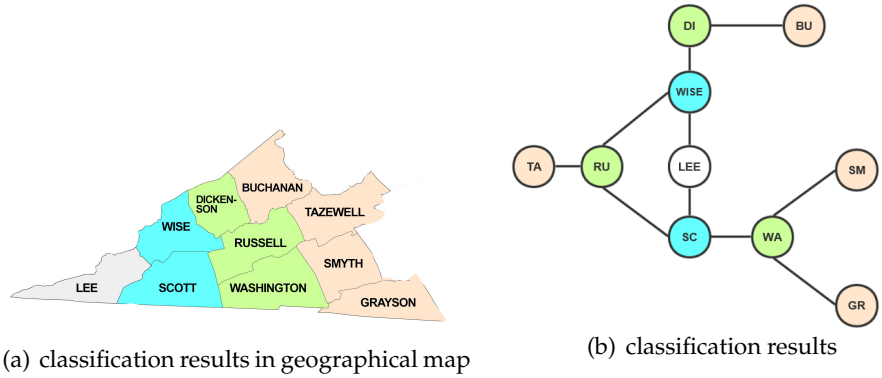
in our model.



(a) Geographical location of the LEE county (b) The model network of the LEE county

Figure 1: An example county and its network

Now we introduce the BSM model in detail. We divide the factors affecting the volume of drug cases in a county into internal and external factors. The internal factors are determined by the number of cases in the current county, and the measurement of external factors is identified by the number of cases involving current county in other counties. In order to describe the number of drug cases in a formulaic way, we give the concept of classification. [9]Assuming that the current county is A, we divide the counties around A into multiple levels in the network. The first-level neighbors are the counties adjacent to A. While the second-level neighbors are the counties adjacent to the first-level neighbors and not adjacent to A, and so on. For example, apply our classification rules for IEE which is illustrated in Figure 1(b), Figure 2(b) shows the results.



(a) classification results in geographical map (b) classification results

Figure 2: An example county and its classification

We believe that the drug cases of a county is affected by itself and surrounding neighbors. Hence, to study the influence of external and internal factors on the number of cases in the county, we established a formula as shown below.

$$g_{t+1}(u) = k_0 + k_1 g_t(u) + k_2 \sum_{v_1 \in D_1} g_t(v_1) + k_3 \sum_{v_2 \in D_2} g_t(v_2) + \ldots + k_{n+1} \sum_{v_n \in D_n} g_t(v_n) \quad (1)$$

In order to simplify the model, we study the distribution of drug cases in five states and counties in 2010. Due to space constraints, we only pick one drug in one state, heroin of VA, as an example to illustrate the study results in detail. The distribution of heroin cases in VA state is shown in Figure 3. Obviously, the darker the color, the

greater the number of cases. By observing Figure 3, we can find that most of the cases are concentrated and county-centered. Therefore, we can assume that the farther the distance is, the less influence the county has on the county. So for a county, we only consider internal factors, first-level neighbors, second-level neighbors, and third-level neighbors to simplify our model, and no longer consider farther counties. The formula for the effects of external and internal factors on drug cases in a county can be simplified as follows.

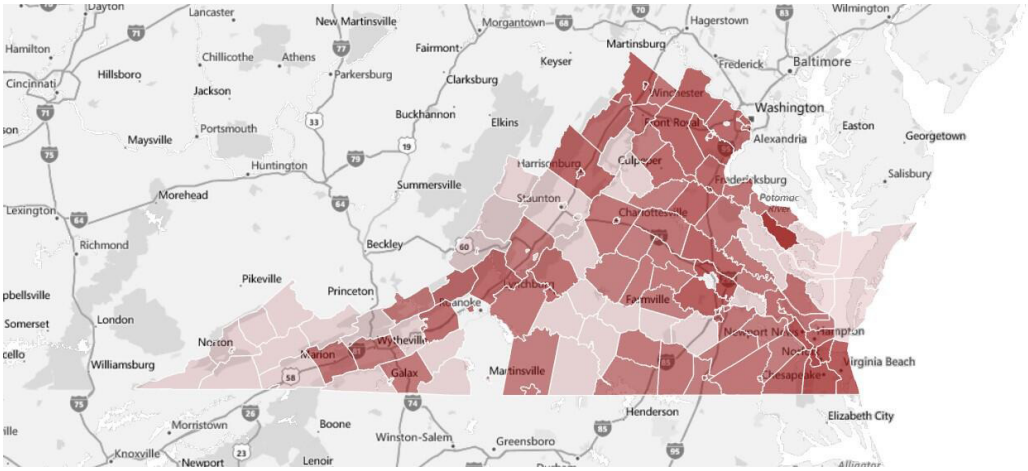


Figure 3: The distribution of drug cases of VA in 2010

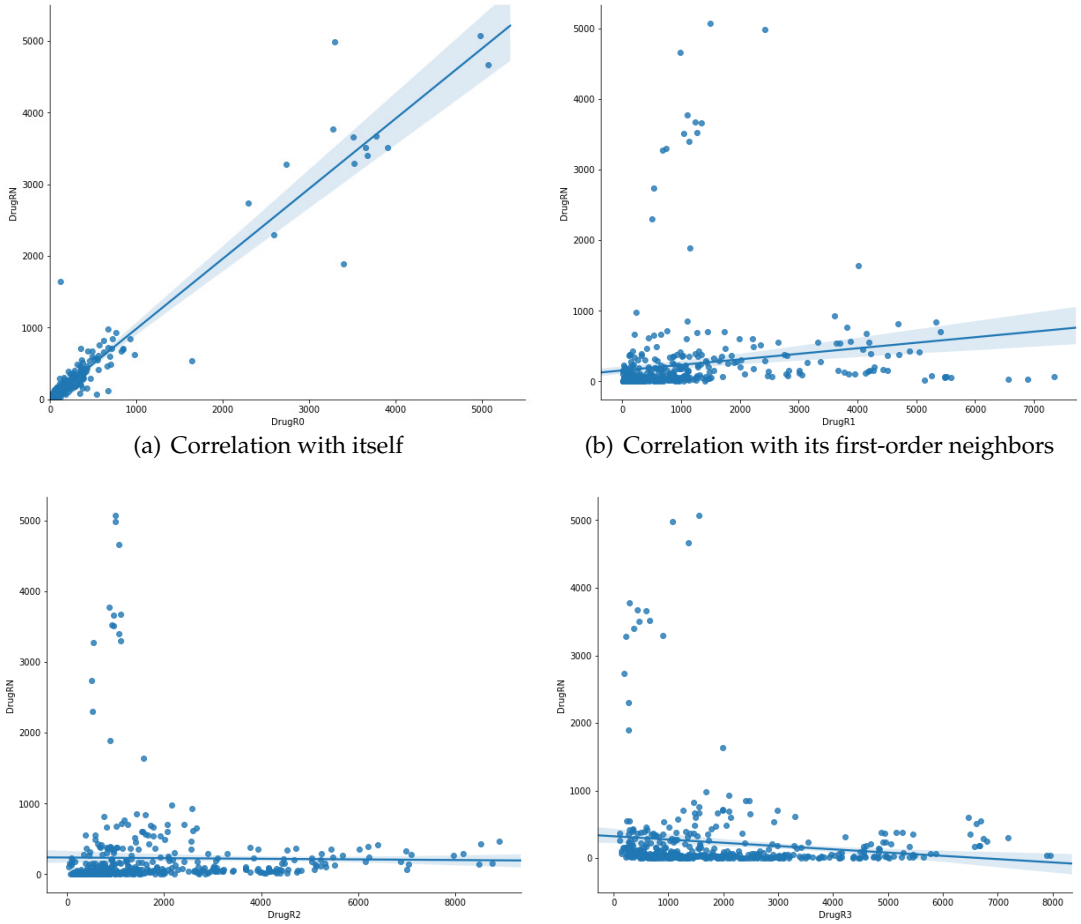
$$g_{t+1}(u) = k_0 + k_1 g_t(u) + k_2 \sum_{v_1 \in D_1} g_t(v_1) + k_3 \sum_{v_2 \in D_2} g_t(v_2) + k_4 \sum_{v_3 \in D_3} g_t(v_3) \tag{2}$$

To find the correlation and trend of each variable and the dependent variable in Equation 2, we analyze the drug cases data from 2010 to 2016. The results is illustrated in Figure 4. We can see that there is a clear linear relationship between the number of cases in the second year and the number of those this year, which roughly distribute around a straight line. The cases in the second year and those in the neighbors are basically in a straight line except for some singular points. That means, for a county, the number of drug cases next year is linearly correlated with the current number of drug cases in itself, first-level neighbors, second-level neighbors and third-level neighbors. Although they all show linear relationships, the extent of their influence is different. With the distance from other counties, the impact is getting lower and lower. For the third-level neighbors, the linear correlation is almost parallel to the X-axis. For higher-level neighbors, it eventually tends to be a straight line parallel to the X-axis. Therefore, it also reflects that it is reasonable for us to study the third order neighbor.

3.2 Linear Regression

Linear regression is a statistical analysis method that uses the regression analysis in mathematical statistics to determine the quantitative relationship between two or more variables. It has been widely used in mathematics, finance, economics and so on. In our BSM model, we employ linear regression to fit and find the appropriate parameters.

To get the appropriate parameters, we employ all data from 2010 to 2017 for 5-fold cross-validation. We divide the data set into five parts, taking four of them as training set



(a) Correlation with itself

(b) Correlation with its first-order neighbors

(c) Correlation with its second-order neighbors

(d) Correlation with its third-order neighbors

Figure 4: Correlation Analysis

and one as testing set in turn. We perform five pieces of training using linear regression, and take the average of each result as the final parameter. The five-fold cross-validation flow chart is shown in Figure 5.

To describe clearly, let take the heroin incidents of VA state as an example, we conducted five pieces of training. The parameters for each training are listed in Table 2. The final parameters are the last row in this table. Then, we use these parameters to simulate the test set.

Now, we put the final parameters into the Equation 2 and simulate the test set. The simulation results are illustrated in Figure 6. The red line represents the predicted outcome of the test set, and the green line represents the actual result. Obviously, the fitting

Table 2: Training Parameter Results

Rounds	k_0	k_1	k_2	k_3	b
(1)	0.7501	0.1853	0.0256	0.0342	4.9764
(2)	0.8359	0.0734	0.0548	0.0258	8.9532
(3)	0.9631	0.1529	0.0679	-0.0127	12.7953
(4)	0.7064	0.1042	0.0159	0.0049	5.6686
(5)	0.8462	0.1162	0.0493	-0.0007	-5.6686
average	0.8203	0.1264	0.0427	0.0103	5.3407

Figure 5: Five-fold cross-validation flow chart

conclusion is satisfactory.

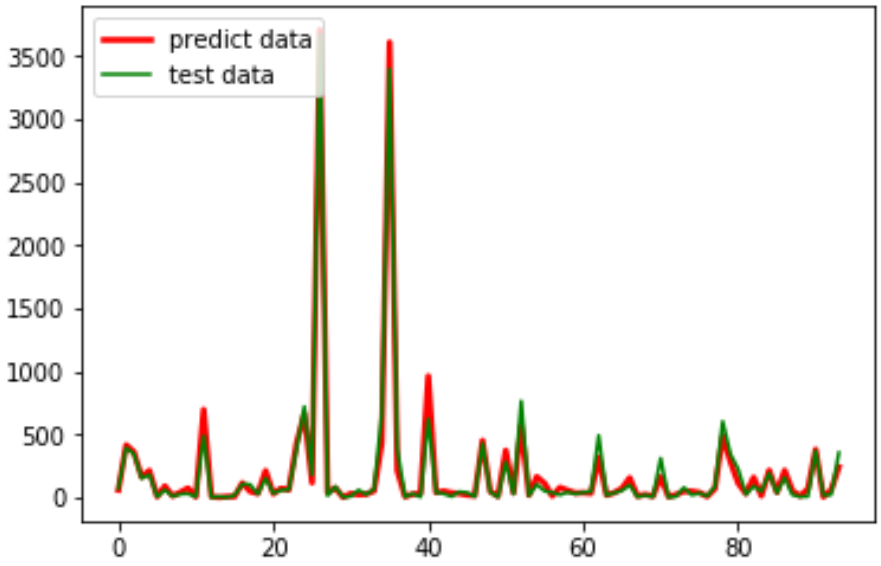


Figure 6: Simulating Results using final parameters

3.3 Characteristics of Spread

Based on the formula parameters obtained by linear regression which is shown in Table 2, it is found that the impact of a county's data this year on the second year reaches 0.9, while the weight of other counties on its second year data is less than 0.1. Therefore, we can conclude that in the process of drug diffusion, a county is the most affected by itself, and is less affected by the proliferation of other counties.

In addition, by observing the parameter relationship of the three-level neighbors, we can find that the farther the distance is, the smaller the parameter is. This finding validates our hypothesis and proves the correctness of the model. Therefore, we can

think that the farther the distance is, the less influence a county has on the other county.

Because of itself influence, we can infer that the spread of drugs has the characteristics of "The rich get richer"; due to the influence of neighbors, we can infer that the transmission of drugs has "The rich bring before the rich" features.

3.4 CountyRank Algorithm

According to the characteristic discovered by our BSM model, namely, "The rich get richer" and "The rich bring before the rich", we have reason to believe that if a county is a location where a drug started to abuse, the number of reported drug cases in this county and counties around this county will be higher. After referring to a large number of node mining algorithms, according to our model, we design an improved PageRank algorithm, named as CountyRank, to calculate the probability that a node is an originating node and trace the source of drug abuse.

In the classic PageRank algorithm [10], the PageRank value represents the importance of the node. The metric refers to two aspects, one is the indegree of the node p_i, and the other is the PageRank value of a node p_j pointing to the node p_i. [4]However, for the spread of drugs, there is no considerable difference in the number of adjacent counties in different counties, and the number of cases in the second year of a county depends more on the number of cases in this year. Hence, we consider the above differences and employ random walks to optimize and get the following formula: [6]

$$CR(p_i) = (1-q)\frac{sum(p_i)}{\sum_{k=1}^{N} sum(p_k)} + q\{\sum_{p_j \in D}(\frac{CR(p_j)}{L(p_j)} * \frac{\beta}{\alpha+\beta}) + CR(p_i) * \frac{\alpha}{\alpha+\beta}\} \quad (3)$$

Now, we write Equation 3 as a matrix:

$$CR = (1-q)\begin{bmatrix} \frac{sum(p_i)}{\sum_{k=1}^{N} sum(p_k)} \\ \vdots \\ \frac{sum(p_N)}{\sum_{k=1}^{N} sum(p_k)} \end{bmatrix} + q\{\frac{\beta}{\alpha+\beta} * \begin{bmatrix} \frac{1}{L(p_1)} & \cdots & \frac{1}{L(p_N)} \\ \vdots & \ddots & \vdots \\ \frac{1}{L(p_1)} & \cdots & \frac{1}{L(p_N)} \end{bmatrix} * M + \frac{\alpha}{\alpha+\beta} I_N\} CR \quad (4)$$

$*$ Multiplies the corresponding positional elements between the matrices.

We define $countyRank0$ and T by Equation 5 and Equation 6.

$$CountyRank0 = \begin{bmatrix} \frac{sum(p_i)}{\sum_{k=1}^{N} sum(p_k)} \\ \vdots \\ \frac{sum(p_N)}{\sum_{k=1}^{N} sum(p_k)} \end{bmatrix} \quad (5)$$

$$T = \frac{\beta}{\alpha+\beta} * \begin{bmatrix} \frac{1}{L(p_1)} & \cdots & \frac{1}{L(p_N)} \\ \vdots & \ddots & \vdots \\ \frac{1}{L(p_1)} & \cdots & \frac{1}{L(p_N)} \end{bmatrix} * M + \frac{\alpha}{\alpha+\beta} I_N \tag{6}$$

According to Equation 5 and Equation 6, we simplify Equation 4 to get Equation 7.

$$CR = (1-q)CountyRank0 + qT * CR \tag{7}$$

In Equation 7, T is a random matrix and it is irreducible and non-periodic. Therefore, for multiple iterations of Equation 7, the CountyRank value will converge to a stable value, which is the probability that a county is the origin of a drug.

We use the above method to calculate the CountyRank value for heroin in each county in VA state and compare it to the number of reported cases in 2010. The comparison results are illustrated in Figure 7. In Figure 7, the depth of the background red represents the number of cases in the county in 2010. The thicker the color, the higher the number of cases. While the brightness of the upper circle represents the CountyRank value of a county. The brighter the ring, the higher the CR value, the more likely the county is an drug abuse origin. It is obviously that where the CR value is high, the number of cases is large, and the CR value in the vicinity is also high.

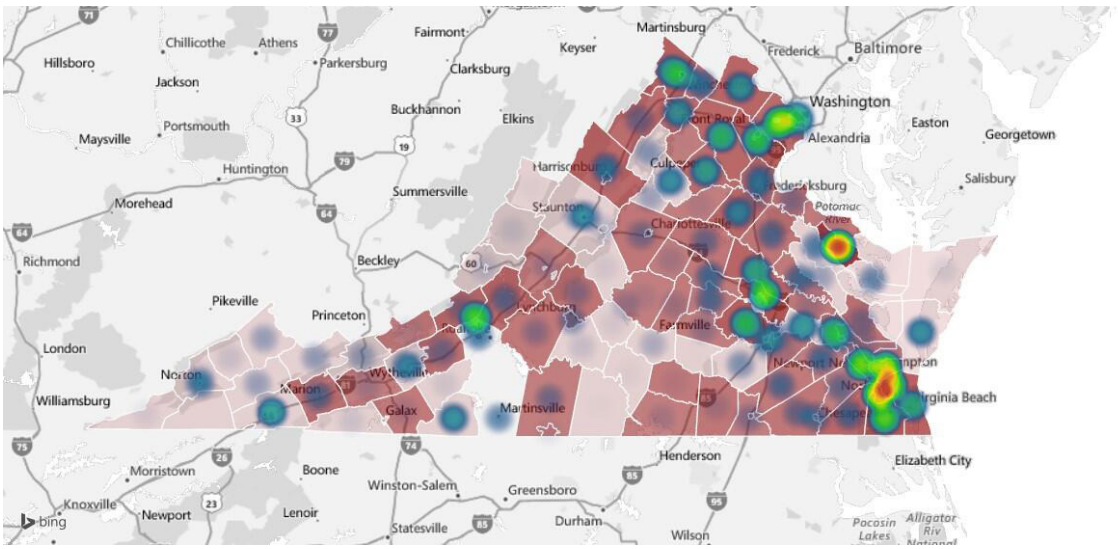


Figure 7: Relevance mining

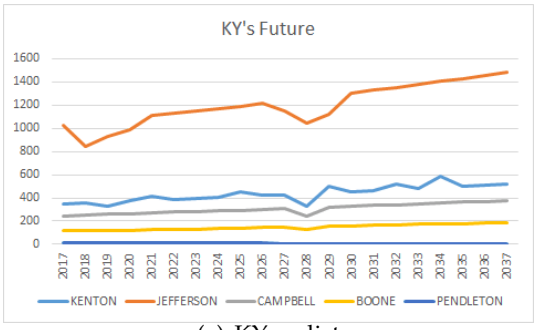
At last, we also find the possible origin of heroin in each of the five states. In each state, we select five counties with the highest CountyRank value as their possible source. The results are shown in Table 3.

3.5 Prediction

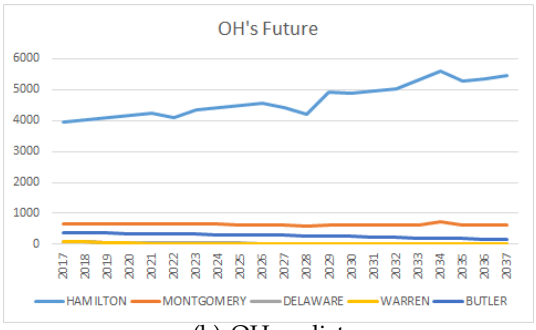
In the case of other factors such as policy are consistent, we predict when and where the spread of drugs will be severe. Due to a large number of counties in the five states, we

Table 3: Possible origins of heroin in five states

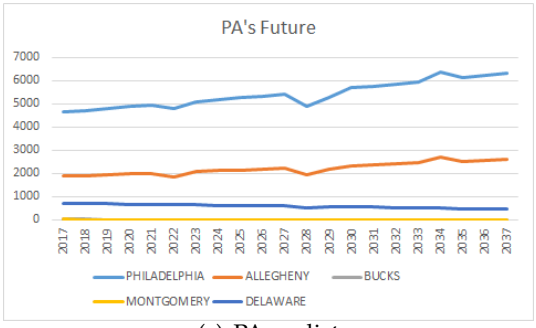
state	origins
KY	KENTON, JEFFERSON, CAMPBELL, BOONE, PENDLETON
OH	HAMILTON, MONTGOMERY, DELAWARE, WARREN, BUTLER
PA	PHILADELPHIA, ALLEGHENY, BUCKS, MONTGOMERY, DELAWARE
VA	RICHMOND, PORTSMOUTH CITY, HAMPTON CITY, NORFOLK CITY, NEWPORT NEWS CITY
WV	MONONGALIA, BERKELEY, MARION, HANCOCK, WETZEL



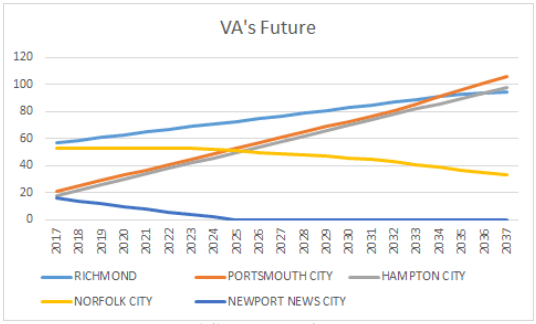
(a) KYpredict



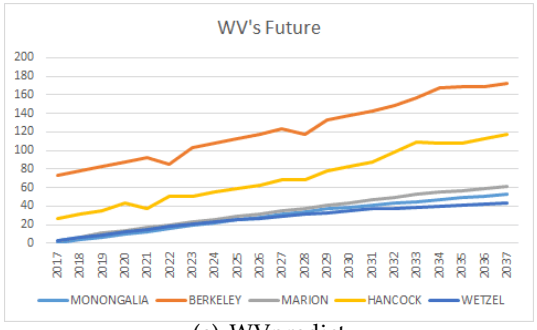
(b) OHpredict



(c) PApredict



(d) VApredict



(e) WVpredict

Figure 8: Predict results of each states

decide to select a part of each state which develops faster and is the most drug-abusing county, as a representative of the state, to predict the trend.

According to our discovery of the law of âĂIJthe rich get richerâĂİ and the CountyRank algorithm, we select five counties with the highest CountyRank value, the most potent county, and forecast these counties to understand the future drug spread trend in the United States.

We forecast the number of drug-related cases at that time for each time node in each

county selected at one-year intervals. Visualize the results of the five states and the results are shown in Figure 8.

For the forecast map of OH in Figure (b), the county with a blue curve is currently a serious drug city, and there is still great potential in the future. The counties in other color curves have flat future drug development and does not need to be too concerned.

We believe that the state that a state needs to pay attention to and intervene depends mainly on the growth rate of the number of cases, so we set a county's threshold to be related to its annual growth.

We summed the number of cases in all of the first-level neighbors of a county last year and divided them by the number of first-level neighbors. After many experiments, we believe that 20% of this value is a safe value. When the county's annual growth is greater than 20% of this value, we think that the county's growth is unusually fast and needs to be taken seriously by the state.

4 The Refined Transmission Model

As described in Section 3, a conclusion can be drawn that the influence of a county's internal factors is exceptionally significant. In fact, the socio-economic factors always affect the internal elements of a county, such as public security and personal quality. In this section, we explore and analyze socio-economic characteristics, try to find out its relationship with drug transmission, and incorporate socio-economic factors into our model, allowing our model to consider more details.

4.1 Correlation Analysis

For socio-economic data sets, we first manually screen out some of the characteristics that may be related to drug spread. Then we calculate the correlation by computing the covariance of each feature and the number of reported drug cases. The higher the absolute value of covariance, the stronger the relationship between this feature and the number of drug cases.

After computing the correlation of each feature and the number of reported drug cases, we sort the correlations according to the absolute values, and organize it into a histogram as shown in Figure 9. We reveal the ten most relevant attributes. Due to the excessive number of eigenvalues, we only select the 20 features with the most relevance for display. As can be seen from the Figure 9, the ten most relevant attributes are unmarried, Widowed, Separation, Couple, marriageable age, divorce, race, etc.

Before we do the ordering of importance, we combine and decompose certain features. For instance, for the characteristics of race, the data set divides the race into a variety of kindsthe German race, the Greek race, the Italian race, and so on. We believe that this division is too detailed, so it is combined according to the geographical distance, and the similarities of the regions are combined into one feature, to obtain a more macro model.

We try a variety of algorithms, such as random forests, K-means, to order the impor-

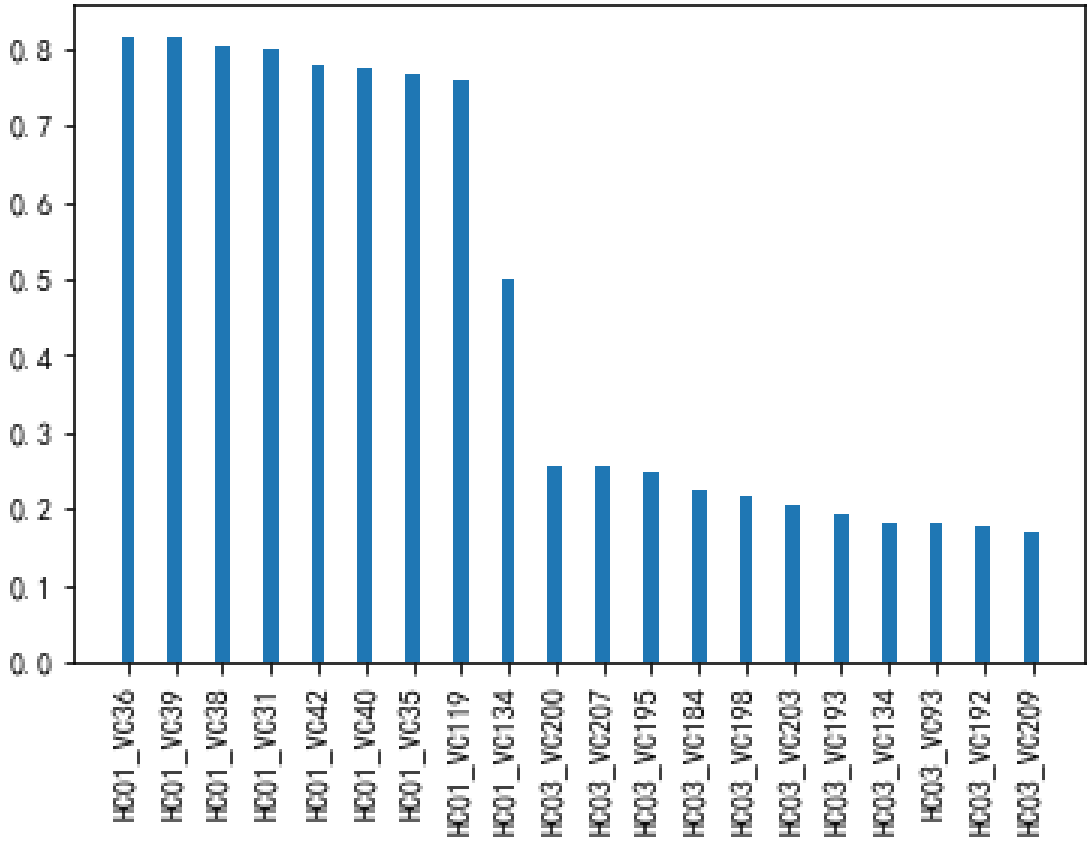


Figure 9: Relevance mining

tance of highly correlated attributes. Then we pass the 5-fold cross-validation to see the accuracy, pick the algorithm with the highest efficiency of results, and trust its feature importance. Finally, we chose the random forest algorithm to compute the importance of features.

4.2 Model with Socio-economic Factors

By the above correlation analysis, we have selected the seven most important attributes, namely, unmarried, widowed, separation, couple, marriageable age, divorce, race. To refine our model, we add these seven attributes as internal factors in the influencing factors and add them to our model and get a new model formula as follows.

$$g_{t+1}(u) = k_0 + k_1 g_t(u) + k_2 \sum_{v_1 \in D_1} g_t(v_1) + k_3 \sum_{v_2 \in D_2} g_t(v_2) + k_4 \sum_{v_3 \in D_3} g_t(v_3) + \lambda_1 f(e_1) + \lambda_2 f(e_2) + \cdots + \lambda_7 f(e_7) \tag{8}$$

Similarly, we use linear regression to find and fit the parameters of Equation 8. We used a five-fold crossover test to increase the accuracy of the parameters. We provided the heroin data for each state. Table 4 is shown the results of parameters computed by Equation 8.

Table 4: Training Parameter of socio-economic fectors

State	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
KY	-0.0103	-0.0036	0.0013	0.0206	-0.0287	-0.0083	0.0056
OH	-0.0384	-0.0013	0.0002	0.1433	0.0061	-0.0244	0.0041
PA	-0.0231	-0.0143	0.0109	-0.0567	0.0172	0.0547	0.0088
VA	0.0084	-0.0012	-0.0018	-0.0047	-0.0006	-0.0001	0.0018
WV	0.0061	-0.0013	-0.0036	0.0095	0.0355	-0.0064	0.0032

In the case of heroin in VA, the following figure (Figure 10) shows a simulation result of the training set after a five-fold cross-check.

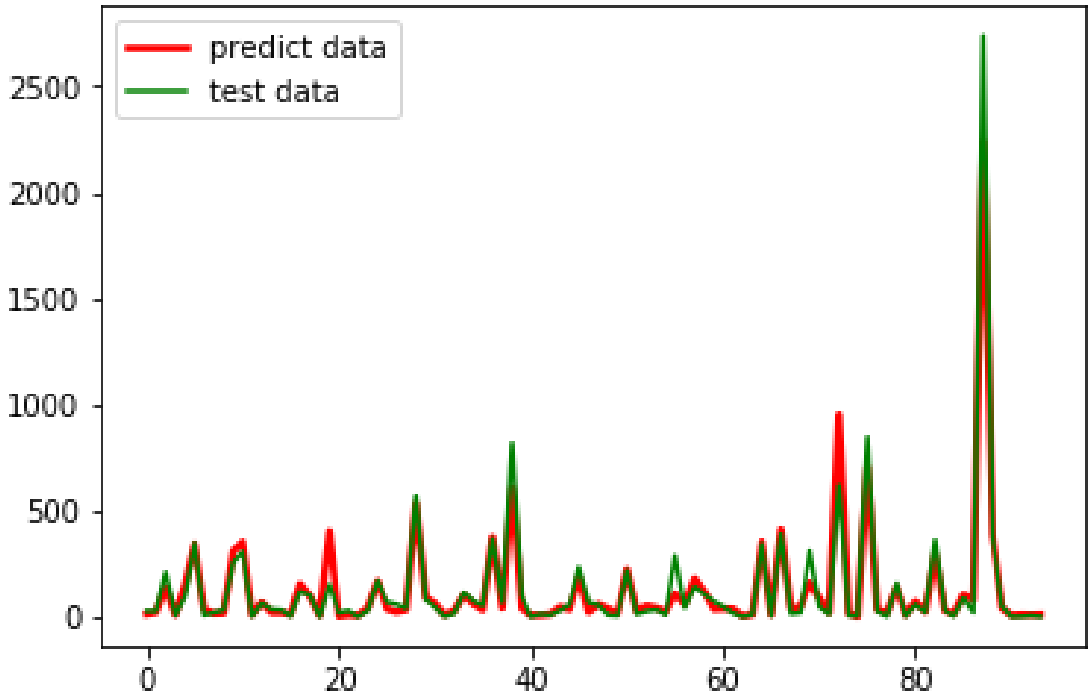


Figure 10: Relevance mining

We also employ the economic model to forecast and simulate the heroin of KENTON and JEFFERSON in KY. Models were used to simulate their trends from 2010 to 2017 and compared with actual data. As can be seen from Figure 11, the fitting results are good.

5 Our Strategy

In response to the influencing factors we found in the socio-economic model, we propose the following solutions:

- Increasing science and publicity to make the public more aware of the harm of opium and other addictive drugs to human beings. Pursuing a short and meaningless pleasure requires a considerable price.

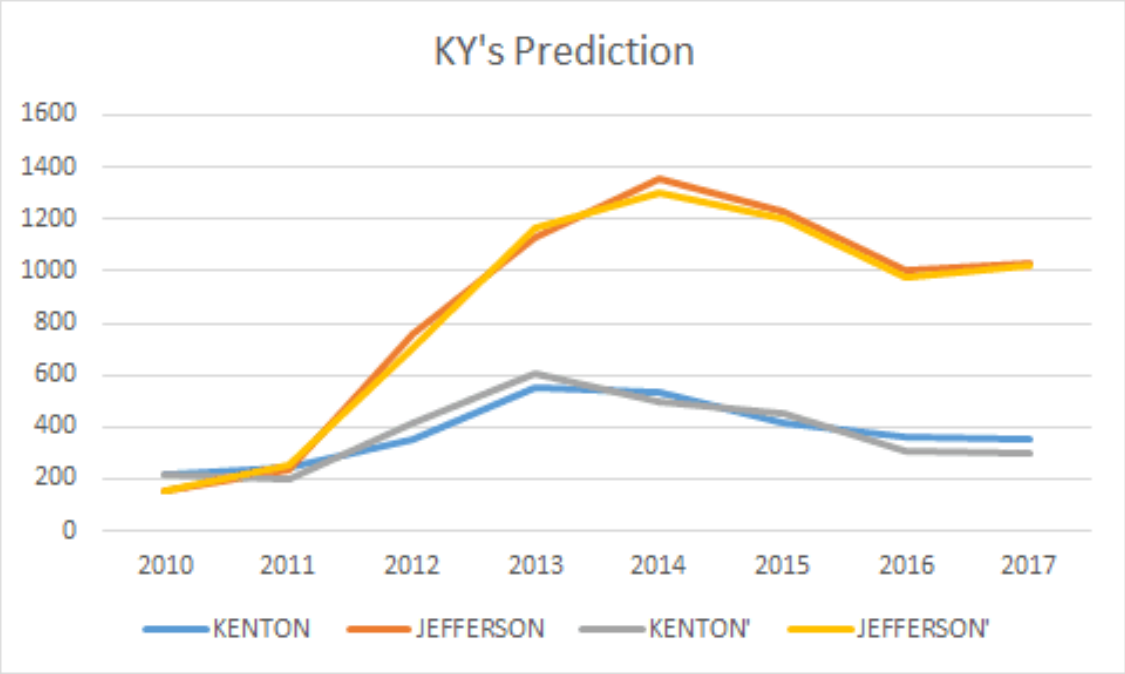


Figure 11: Heroin in KY using socio-economic model

- Strengthen the psychological counseling of the people, try to give emotional care, improve the quality of community psychologists, and provide more professional guidance and advice to those in need.
- Raise attention to the degree of social stability. When the number of unemployed and unemployed in society is high or is increasing rapidly, relevant government departments should strengthen the supervision of addictive drugs and alleviate a large number of unemployed.
- Equality of ethnicity and increased oversight in places where particular ethnic groups gather.

The above strategies aim at marriage and ethnicity. Therefore, we apply the effects on marriage and ethnicity to the model, apply it to Heroin transmission in KY, and use the model to simulate the results of our strategy. The results are illustrated in Figure 12 and Figure 13. We can see that the rate of drug spread significantly reduces.

6 Strengths and weaknesses

6.1 Strengths

- **Comprehensive.** When analyzing the drug transmission in each state, we divide the influencing factors into internal and external factors. The internal factors consider the race, education, number of disabled people, marital status, etc. In the external factors, we think the distance according to whether it is adjacent. Not only do we comprehensively consider indicators [7], but we also find the quantitative relationship between them.

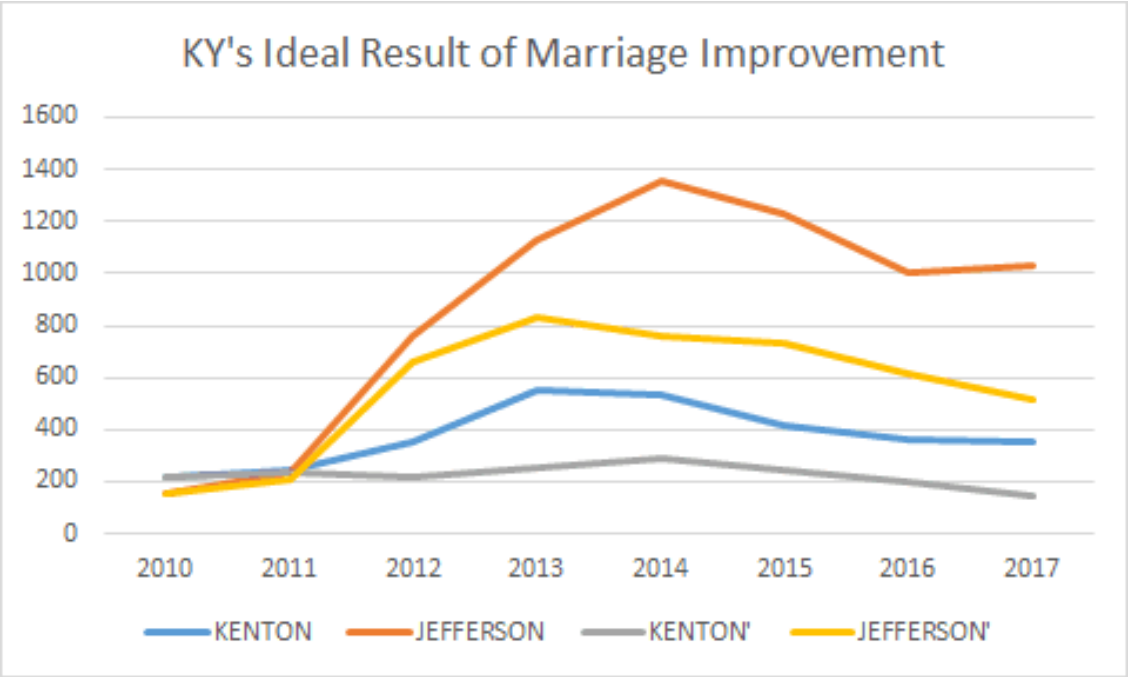


Figure 12: The marriage factor in KY

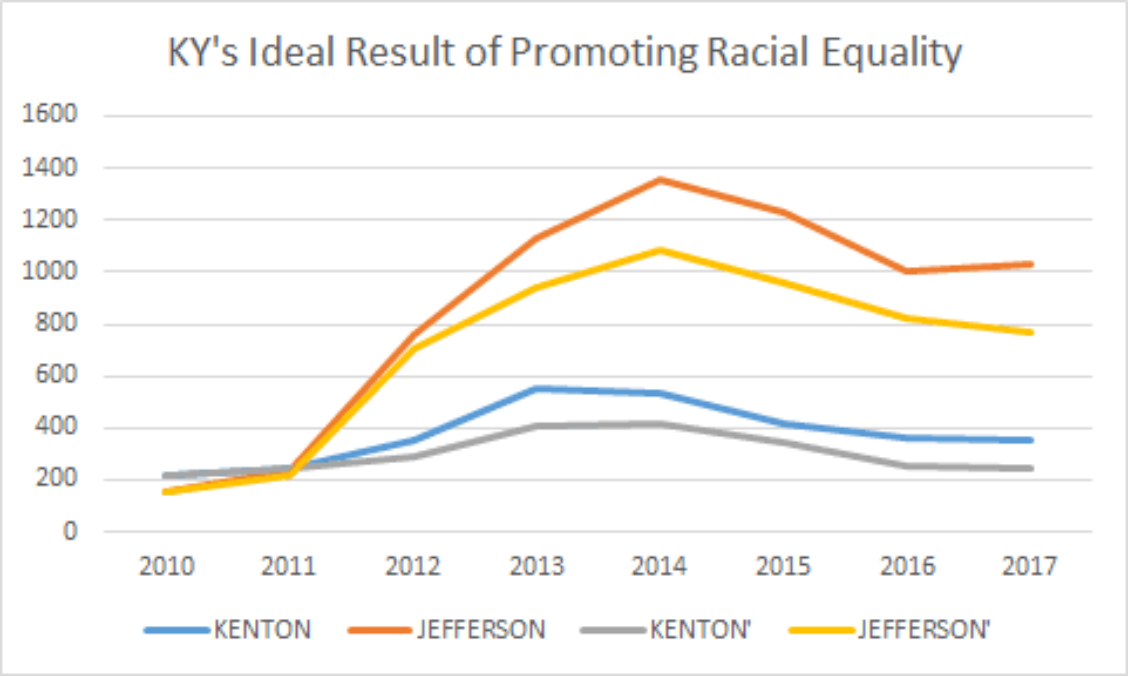


Figure 13: The race factor in KY

- **Efficient.** In the process of determining the parameters of the graph propagation model, we used the linear regression method to make the metrics more realistic. We find the most suitable settings more efficiently.
- **High versatility.** Since the data set determines our parameters, different data sets can get different suitable formulas and parameters, which allows our model to be applied to other regions more widely.
- **Low Feature and Evaluation Complexity.** We refine seven features out of total 600 ones. This makes our model better understand.

6.2 Weakness

- **No errors were considered.** In the socio-economic data, we only consider the estimates without considering the impact of the errors.
- **We simplify the distance factor.** We streamline the distance to whether it is adjacent, and do not consider the weight of the edges in the network. This simplification may make our model think that the influence of counties in the different distance is the same.
- **Lack of possible costs.** We do not consider the labor and economic costs of the government to solve the problem.

7 Conclusion

In this paper, we explore the factors related to drug transmission and the characteristics of spreading through the analysis and research of reported drug cases. Besides, we propose three practical actions for the Chief Administrator, DEA/NFLIS Database to solve these problems. Here we conclude our findings. Moreover, we recommend three possible activities for the Chief Administrator, DEA/NFLIS Database to solve these problems. Here we find our conclusions.

First, in order to understand the laws of drug transmission, we have established a graph propagation model and improved it. We considered the external factors of the influence of surrounding counties and explored the internal factors of marriage, race and drug transmission. [1]

Second, We traced the source of heroin drug transmission and found some rules about the spread of drugs. We see that drug transmission is mainly affected by internal factors, and the spread has the law of âĂIJThe rich get richer.âĂİ We also found that drug transmission is primarily influenced by marriage and ethnic factors.

At last, we proposed four solutions to solve the problem: increase the science of drugs, strengthen the psychological counseling of the people, raise the concern for social stability, and strengthen supervision to make people equal.

References

[1] Basu, S., Andrews, J.R., Poolman, E.M., Gandhi, N.R., Shah, N.S., Moll, A., Moodley, P., Galvani, A.P., Friedland, G.H.: Prevention of nosocomial transmission of extensively drug-resistant tuberculosis in rural south african district hospitals: an epidemiological modelling study. The Lancet **370**(9597), 1500–1507 (2007)

[2] Blower, S.M., Dowlatabadi, H.: Sensitivity and uncertainty analysis of complex models of disease transmission: an hiv model, as an example. International Statistical Review/Revue Internationale de Statistique pp. 229–243 (1994)

[3] Dechter, R., Kask, K., Mateescu, R.: Iterative join-graph propagation. In: Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence. pp. 128–136. Morgan Kaufmann Publishers Inc. (2002)

[4] Desikan, P., Pathak, N., Srivastava, J., Kumar, V.: Incremental page rank computation on evolving graphs. In: Special interest tracks and posters of the 14th international conference on World Wide Web. pp. 1094–1095. ACM (2005)

[5] Hudgens, M.G., Longini Jr, I.M., Vanichseni, S., Hu, D.J., Kitayaporn, D., Mock, P.A., Halloran, M.E., Satten, G.A., Choopanya, K., Mastro, T.D.: Subtype-specific transmission probabilities for human immunodeficiency virus type 1 among injecting drug users in bangkok, thailand. American journal of Epidemiology **155**(2), 159–168 (2002)

[6] Langville, A.N., Meyer, C.D.: Deeper inside pagerank. Internet Mathematics **1**(3), 335–380 (2004)

[7] Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N.: Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 420–429. ACM (2007)

[8] Li, J., Ma, M.: The analysis of a drug transmission model with family education and public health education. Infectious Disease Modelling **3**, 74–84 (2018)

[9] Mateescu, R., Kask, K., Gogate, V., Dechter, R.: Join-graph propagation algorithms. Journal of Artificial Intelligence Research **37**, 279–328 (2010)

[10] Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Tech. rep., Stanford InfoLab (1999)

[11] Saidel, T.J., Des Jarlais, D., Peerapatanapokin, W., Dorabjee, J., Singh, S., Brown, T.: Potential impact of hiv among idus on heterosexual transmission in asian settings: scenarios from the asian epidemic model. International journal of drug policy **14**(1), 63–74 (2003)

Appendices

Here are programmes we used in our model as follow.

Prepare for Linear Regression

```
#-*- coding:utf-8 -*-

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from pandas import DataFrame,Series

from sklearn.cross_validation import train_test_split

from sklearn.linear_model import LinearRegression

data = pd.read_csv("C://Users/zhangjianing/Documents/Tencent Files/2813295041/FileRecv/PAherior

data.drop(['YYYY'], axis=1, inplace=True)

data.drop(['FIPS_Combined'], axis=1, inplace=True)

examDf = DataFrame(data)

new$_$examDf = examDf.ix[:,0:]

print(new_examDf.describe())

print(new_examDf[new_examDf.isnull()==True].count())

print(new_examDf.corr())

sns.pairplot(data, x_vars=['DrugR0'], y_vars='DrugRN', size=7, aspect=1.2, kind='reg')

plt.xlim((0, 5500))

plt.ylim((0, 5500))

plt.show()

sns.pairplot(data, x_vars=['DrugR1'], y_vars='DrugRN', size=7, aspect=1.2, kind='reg')

plt.show()

sns.pairplot(data, x_vars=['DrugR2'], y_vars='DrugRN', size=7, aspect=1.2, kind='reg')

plt.show()

sns.pairplot(data, x_vars=['DrugR3'], y_vars='DrugRN', size=7, aspect=1.2, kind='reg')

plt.show()
```

Linear Regression

```
data=load('.CSV');

g=data(:,3);

g1=data(:,4);

g2=data(:,5);

g3=data(:,6);

g4=data(:,7);

g5=data(:,8);

g6=data(:,9);

g7=data(:,10);

g8=data(:,11);

g9=data(:,12);

g10=data(:,13);

y=data(:,14);

one=ones(length(y),1);

X=[one,g,g1,g2,g3,g4,g5,g6,g7,g8,g9,g10];

Y=y;

[b,bint,r,rint,stats]=regress(Y,X);

b,bint,stats,(stats(4))^0.5
```

CountyRank

```
data=load('.CSV');

lie=data(:,1);

len=length(lie);

q=0.85;

sum=0;

for i=1:len

    sum=sum+data(i,3);

end

R0=data(:,3)/sum;

linjie=load('.CSV');

linjie=linjie(2:len+1,2:len+1);

a=0.8;
```

```
b=0.2;
T=zeros(len);
for i=1:len
    tmp=0;
    for j=1:len
        tmp=tmp+linjie(j,i);
    end
    for j=1:len
        T(j,i)=b/(tmp*(a+b));
    end
end
T=T.*linjie;
for i=1:len
    T(i,i)=a/(a+b);
end
R=R0;
for i=1:100000
    R=(1-q)*R0+q*T*R;
end
R
```