

# A Network Tour of Inter-County Migration in the United States

Final Project Presentation - Team 22

*Course: A Network Tour Of Data Science*

*Students: Fatima Ezzahra, Xiaoyan Zou, Paul Mansat, Anshul Toshniwal*

**EPFL**

I. Motivation and objectives

II. Data Acquisition

III. Data Exploration

IV. Data Exploitation

# Motivation and objectives

3

---

## Objective

Predict the result of the 2016 Presidential Elections at the county level according to migration data

# Motivation and objectives

4

---

Why migration can influence an election ?



More and more  
migration



Migration is a key  
subject in politics

# Motivation and objectives

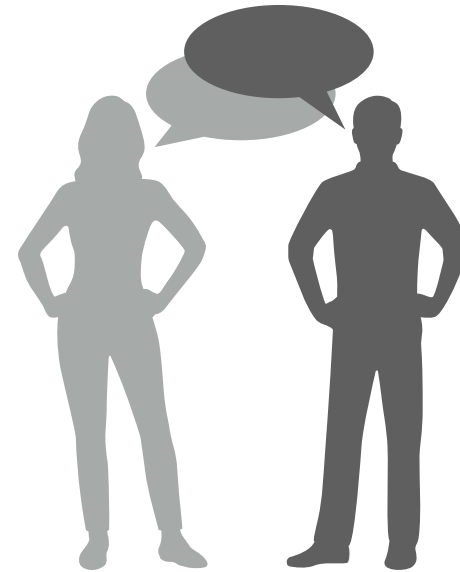
4

---

Why migration can influence an election ?



More and more  
migration



Migration is a key  
subject in politics

# Motivation and objectives

---

5

## Background

### County:

- geographical subdivision of a US states
- 3007 counties in 2019
- define uniquely by FIPS

### Republicans vs Democrats:

- main political forces in the US
- in 2016, 82% of counties were Republican, 28% were Democrat

Republican

Democrats

# Motivation and objectives

---

5

## Background

### County:

- geographical subdivision of a US states
- 3007 counties in 2019
- define uniquely by FIPS

### Republicans vs Democrats:

- main political forces in the US
- in 2016, 82% of counties were Republican, 28% were Democrat

Republican

Democrats

# Motivation and objectives

---

5

## Background

### County:

- geographical subdivision of a US states
- 3007 counties in 2019
- define uniquely by FIPS

### Republicans vs Democrats:

- main political forces in the US
- in 2016, 82% of counties were Republican, 28% were Democrat

Republican

Democrats



# Motivation and objectives

---

5

## Background

### County:

- geographical subdivision of a US states
- 3007 counties in 2019
- define uniquely by FIPS

### Republicans vs Democrats:

- main political forces in the US
- in 2016, 82% of counties were Republican, 28% were Democrat

Republican

Democrats

# Motivation and objectives

---

5

## Background

### County:

- geographical subdivision of a US states
- 3007 counties in 2019
- define uniquely by FIPS

### Republicans vs Democrats:

- main political forces in the US
- in 2016, 82% of counties were Republican, 28% were Democrat

Republican

Democrats

# Motivation and objectives

---

5

## Background

### County:

- geographical subdivision of a US states
- 3007 counties in 2019
- define uniquely by FIPS

### Republicans vs Democrats:

- main political forces in the US
- in 2016, 82% of counties were Republican, 28% were Democrat

Republican

Democrats

# Motivation and objectives

---

5

## Background

### County:

- geographical subdivision of a US states
- 3007 counties in 2019
- define uniquely by FIPS

### Republicans vs Democrats:

- main political forces in the US
- in 2016, 82% of counties were Republican, 28% were Democrat

Republican

Democrats

I. Motivation and objectives

II. Data Acquisition

III. Data Exploration

IV. Data Exploitation

# Data Acquisition

7

## Source and pre-processing of the data

### Migration Data:

From IRS database. Extract two kind of tables:

1.	FIPS origin	FIPS dest.	Returns	Exempt	AGI
	...	...	...	...	...
2.	FIPS	Total mig.	US. mig	Inter. mig.	Ret./Exem.
	...	...	...	...	...

### Election data:

From Guardian website

FIPS	Dem. pct.	Rep. pct.
...	...	...

# Data Acquisition

7

## Source and pre-processing of the data

### Migration Data:

From IRS database. Extract two kind of tables:

1.	<b>FIPS origin</b>	<b>FIPS dest.</b>	<b>Returns</b>	<b>Exempt</b>	<b>AGI</b>
	...	...	...	...	...
2.	<b>FIPS</b>	<b>Total mig.</b>	<b>US. mig</b>	<b>Inter. mig.</b>	<b>Ret./Exem.</b>
	...	...	...	...	...

### Election data:

From Guardian website

<b>FIPS</b>	<b>Dem. pct.</b>	<b>Rep. pct.</b>
...	...	...

# Data Acquisition

7

## Source and pre-processing of the data

### Migration Data:

From IRS database. Extract two kind of tables:

1.	<b>FIPS origin</b>	<b>FIPS dest.</b>	<b>Returns</b>	<b>Exempt</b>	<b>AGI</b>
	...	...	...	...	...
2.	<b>FIPS</b>	<b>Total mig.</b>	<b>US. mig</b>	<b>Inter. mig.</b>	<b>Ret./Exem.</b>
	...	...	...	...	...

### Election data:

From Guardian website

<b>FIPS</b>	<b>Dem. pct.</b>	<b>Rep. pct.</b>
...	...	...



# Data Acquisition

7

## Source and pre-processing of the data

### Migration Data:

From IRS database. Extract two kind of tables:

1.	FIPS origin	FIPS dest.	Returns	Exempt	AGI
	...	...	...	...	...
2.	FIPS	Total mig.	US. mig	Inter. mig.	Ret./Exem.
	...	...	...	...	...

### Election data:

From Guardian website

FIPS	Dem. pct.	Rep. pct.
...	...	...

I. Motivation and objectives

II. Data Acquisition

III. Data Exploration

IV. Data Exploitation

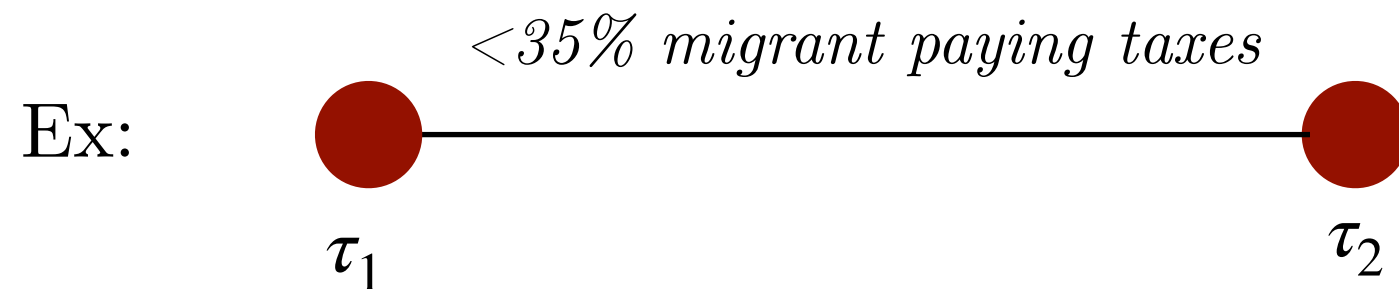
---

## Exempt and return graph

Unweighted graph

Nodes are the county

Edge between two counties  $\tau_1$  and  $\tau_2$  if migration between counties and based on percent of migrant exempt from taxes or filling returns

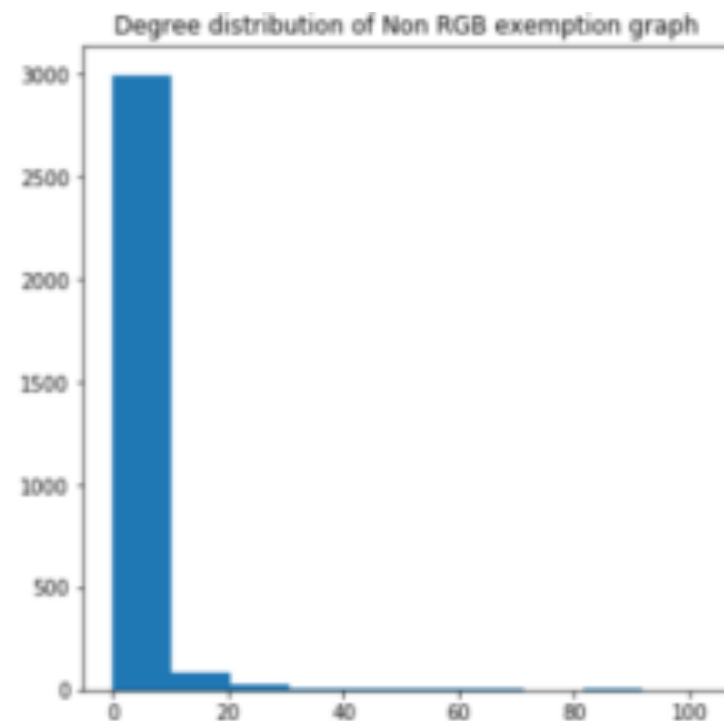


# Data Exploration

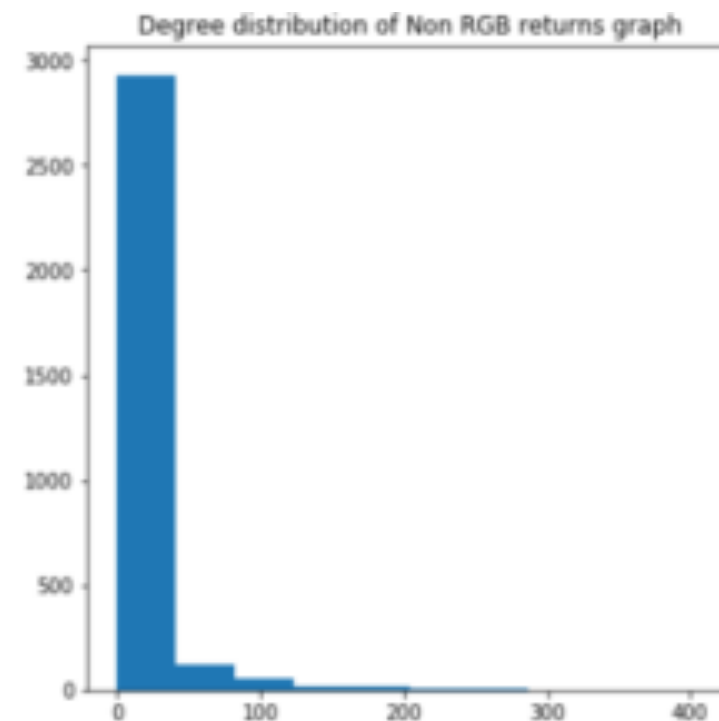
10

## Exempt and return graph

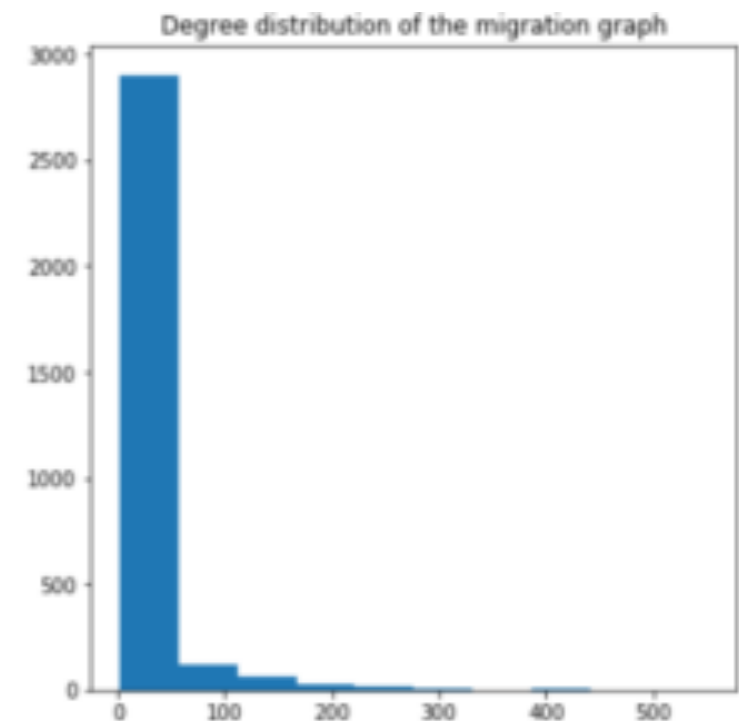
Graphs are scale free:



*Figure 1: >70% migrant paying taxes*



*Figure 2: >35% migrant filling returns*



*Figure 3: total migration graph*

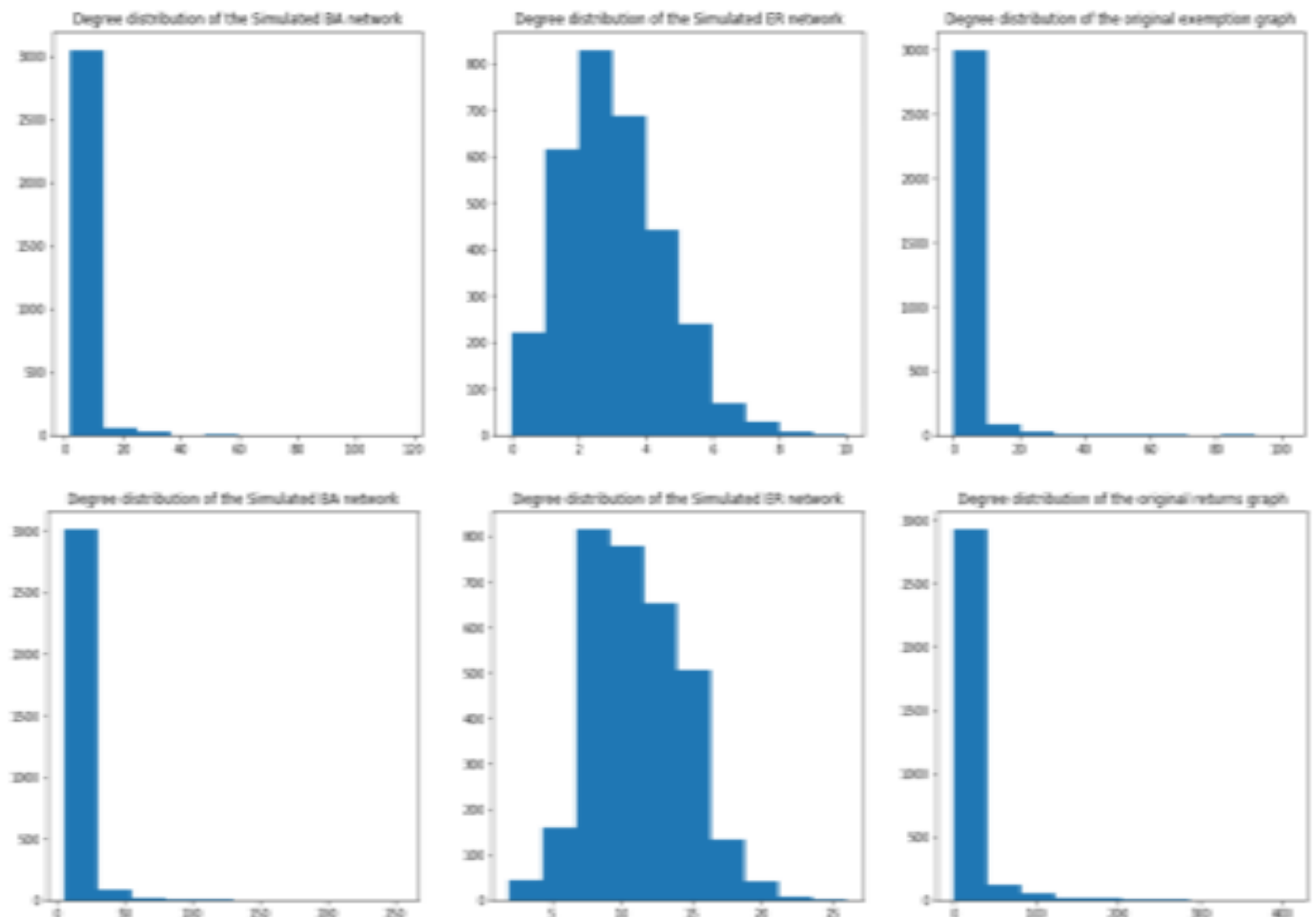
# Data Exploration

11

## Properties of the graph

Exemption graph sparser than the return graph and has lower clustering coefficient

**Exemption and return** graph are best simulated with BA network



I. Motivation and objectives

II. Data Exploration

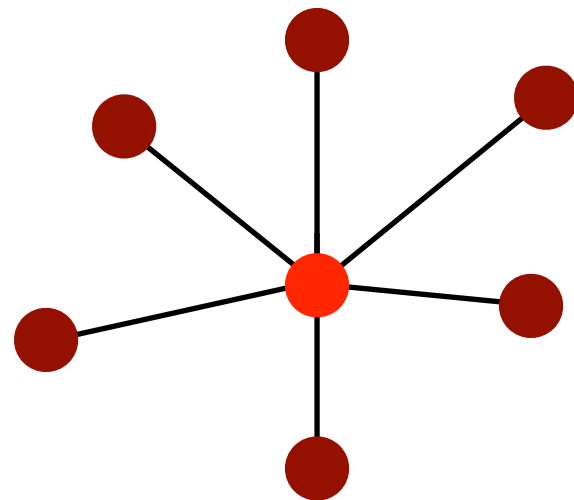
III. Data Acquisition

IV. Data Exploration

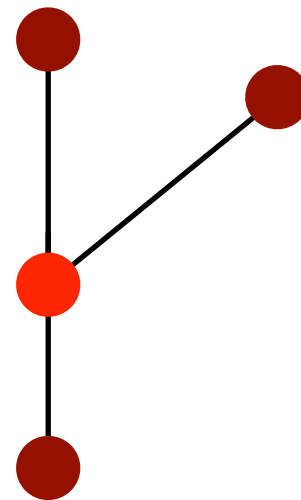
---

## First Attempt: prediction according to degree

Predict according to degree of node:



Democrat



Republican

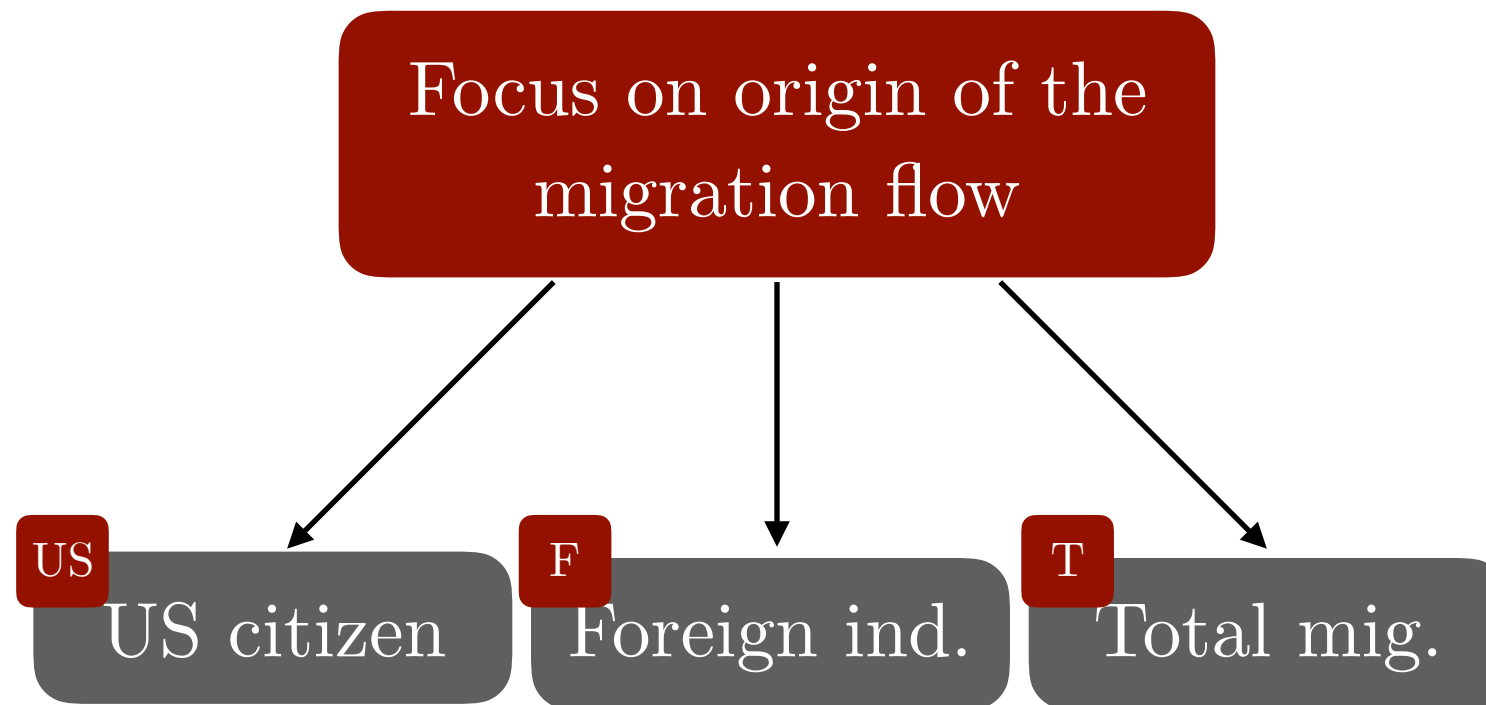
Predict according to degree of average degree of neighbours of node

# Data Exploitation

14

## Second Attempt: Graph Construction

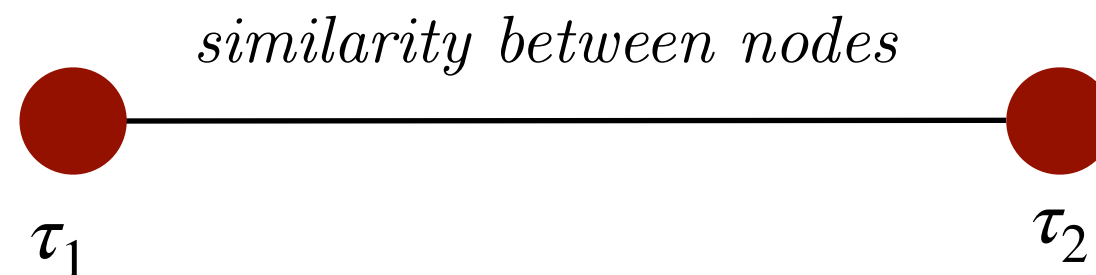
Creation of 3 similarity graph:



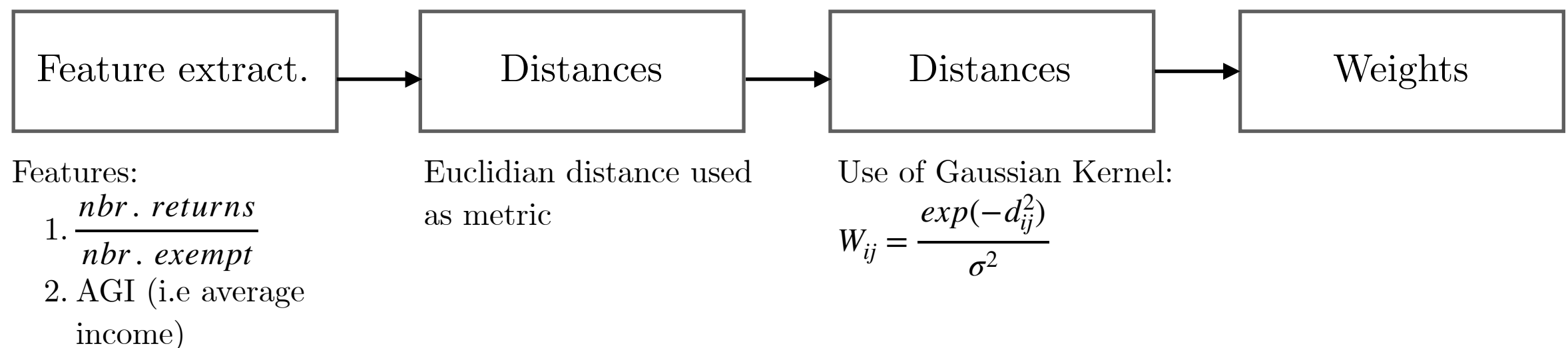


## Second Attempt: Structure of the Graph

**Structure:** 3 weighted adjacency matrix, nodes are the counties



## Building process:

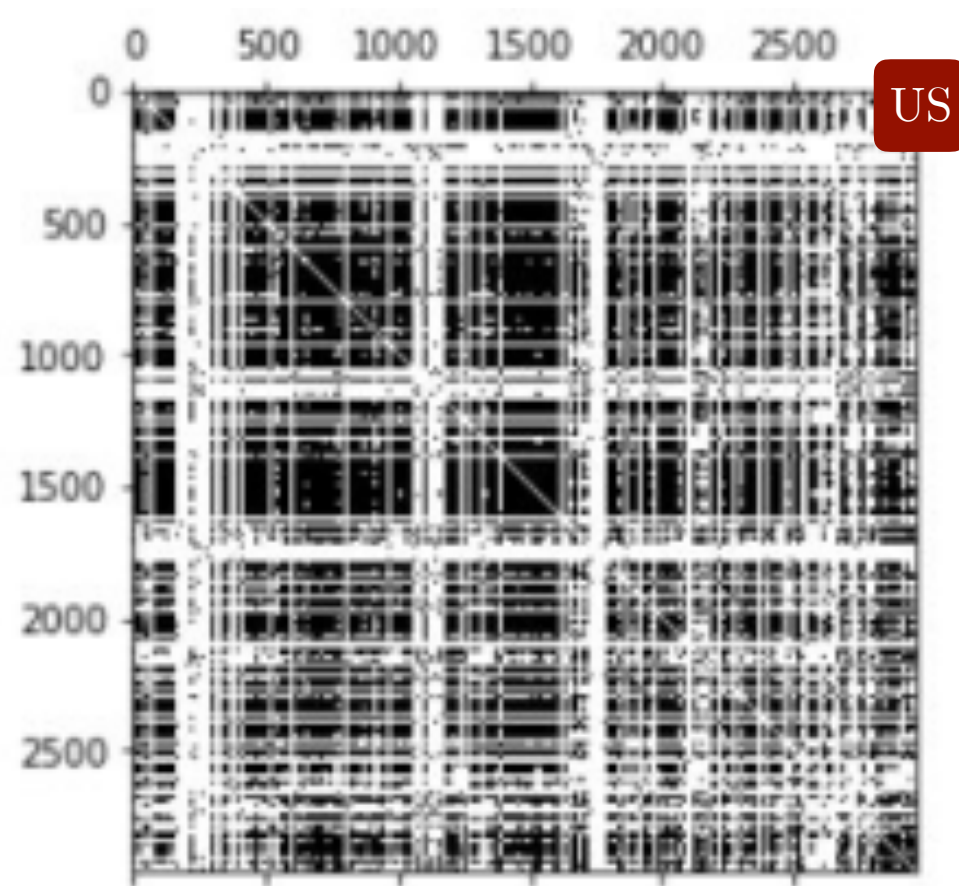


# Data Exploitation

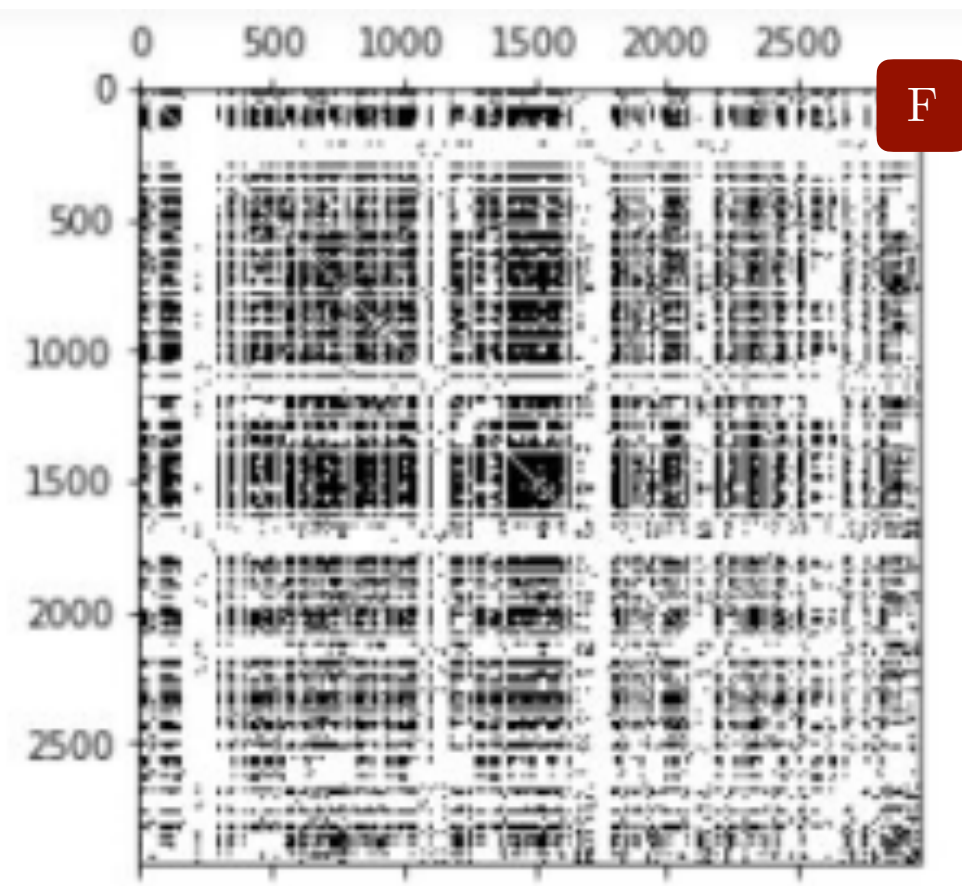
16

## Second Attempt: Adjacency Matrix

Kernel width:  $\mathbb{E}[d_{ij}] \cdot 0.1$



*Figure 1: Adjacency matrix of the total migration flow*



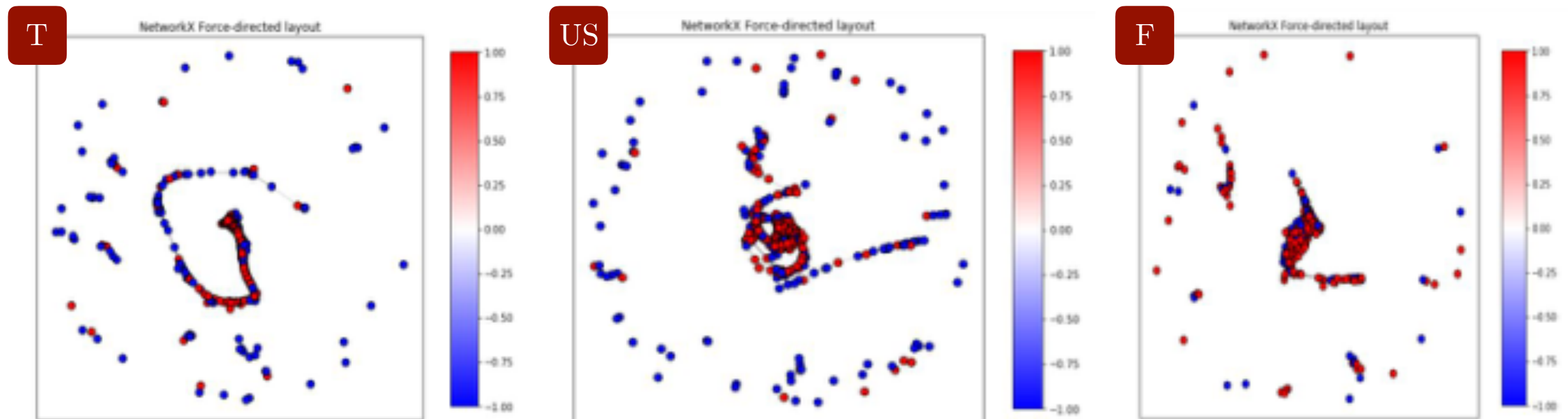
*Figure 2: Adjacency matrix of the US citizen migration flow*

# Data Exploitation

17

## Second Attempt: NetworkX force layout for the similarity graph

Plotting the graph using spring layout simultaneously with the election result of each node:



*Figure 1: Total migration*

*Figure 2: US citizen migration*

*Figure 3: foreigner migration*

# Data Exploitation

18

## Second Attempt: Fourier Analysis

### Process:

Masked winner

GFT

Low Pass  
Filter

iGFT

Filtered winner

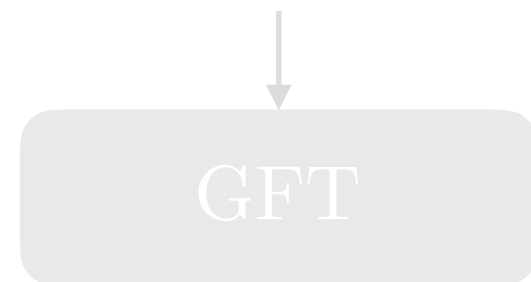
### Results:

county_id	agi	prop_ret/exempt
1.0	-0.13	0.43
2.0	0.45	0.50
3.0	-0.28	0.47
4.0	-0.28	0.43
5.0	-0.23	0.46
6.0	-0.30	0.45
7.0	-0.29	0.50
8.0	-0.16	0.50
9.0	-0.26	0.49
10.0	-0.27	0.46

## Second Attempt: Fourier Analysis

### Process:

Masked winner



Low Pass  
Filter

iGFT

Filtered winner

### Results:

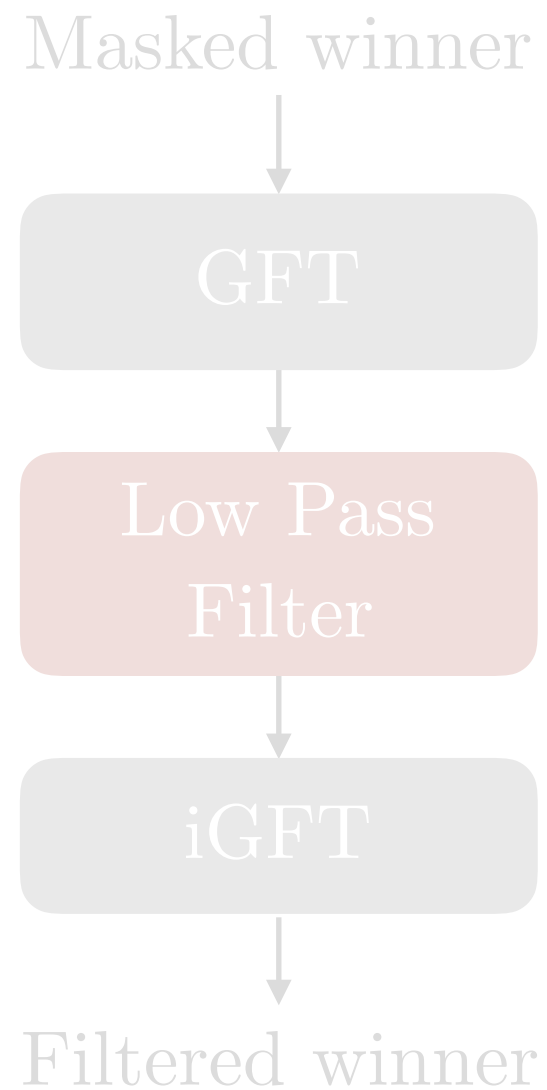
county_id	agi	prop_ret/exempt	winner
1.0	-0.13	0.43	1.0
2.0	0.45	0.50	1.0
3.0	-0.28	0.47	1.0
4.0	-0.28	0.43	1.0
5.0	-0.23	0.46	-1.0
6.0	-0.30	0.45	1.0
7.0	-0.29	0.50	1.0
8.0	-0.16	0.50	1.0
9.0	-0.26	0.49	1.0
10.0	-0.27	0.46	-1.0

# Data Exploitation

18

## Second Attempt: Fourier Analysis

### Process:



### Results:

county_id	agi	prop_ret/exempt	winner	masked_winner
1.0	-0.13	0.43	1.0	1.0
2.0	0.45	0.50	1.0	1.0
3.0	-0.28	0.47	1.0	1.0
4.0	-0.28	0.43	1.0	1.0
5.0	-0.23	0.46	-1.0	0.0
6.0	-0.30	0.45	1.0	1.0
7.0	-0.29	0.50	1.0	0.0
8.0	-0.16	0.50	1.0	1.0
9.0	-0.26	0.49	1.0	1.0
10.0	-0.27	0.46	-1.0	-1.0

## Second Attempt: Fourier Analysis

### Process:

Masked winner

GFT

Low Pass  
Filter

iGFT

Filtered winner

### Results:

county_id	agi	prop_ret/exempt	winner	masked_winner
1.0	-0.13	0.43	1.0	1.0
2.0	0.45	0.50	1.0	1.0
3.0	-0.28	0.47	1.0	1.0
4.0	-0.28	0.43	1.0	1.0
5.0	-0.23	0.46	-1.0	0.0
6.0	-0.30	0.45	1.0	1.0
7.0	-0.29	0.50	1.0	0.0
8.0	-0.16	0.50	1.0	1.0
9.0	-0.26	0.49	1.0	1.0
10.0	-0.27	0.46	-1.0	-1.0

## Second Attempt: Fourier Analysis

### Process:

Masked winner

GFT

Low Pass  
Filter

iGFT

Filtered winner

### Results:

county_id	agi	prop_ret/exempt	winner	masked_winner	filtered_winner
1.0	-0.13	0.43	1.0	1.0	0.29
2.0	0.45	0.50	1.0	1.0	0.36
3.0	-0.28	0.47	1.0	1.0	0.78
4.0	-0.28	0.43	1.0	1.0	0.67
5.0	-0.23	0.46	-1.0	0.0	-0.12
6.0	-0.30	0.45	1.0	1.0	0.69
7.0	-0.29	0.50	1.0	0.0	0.78
8.0	-0.16	0.50	1.0	1.0	0.44
9.0	-0.26	0.49	1.0	1.0	0.78
10.0	-0.27	0.46	-1.0	-1.0	-0.76



## Second Attempt: Fourier Analysis

### Process:

Masked winner

GFT

Low Pass  
Filter

iGFT

Filtered winner

### Results:

county_id	agi	prop_ret/exempt	winner	masked_winner	filtered_winner	prediction
1.0	-0.13	0.43	1.0	1.0	0.29	1.0
2.0	0.45	0.50	1.0	1.0	0.36	1.0
3.0	-0.28	0.47	1.0	1.0	0.78	1.0
4.0	-0.28	0.43	1.0	1.0	0.67	1.0
5.0	-0.23	0.46	-1.0	0.0	-0.12	-1.0
6.0	-0.30	0.45	1.0	1.0	0.69	1.0
7.0	-0.29	0.50	1.0	0.0	0.78	1.0
8.0	-0.16	0.50	1.0	1.0	0.44	1.0
9.0	-0.26	0.49	1.0	1.0	0.78	1.0
10.0	-0.27	0.46	-1.0	-1.0	-0.76	-1.0

## Second Attempt: Fourier Results

Graph	Low pass filter		Heat kernel	
	Mean	Variance	Mean	Variance
Return	0.89	0.008	0.88	0.010
Exempt	0.91	0.007	0.95	0.007
Total_migration	0.93	0.006	0.93	0.009
US_citizen	0.92	0.007	0.93	0.009
Foreigner_individuals	0.85	0.030	0.83	0.020

*Figure 1: Results of Fourier's analysis*

## Third Attempt: Graph convolutional Network

Applied on the return & exempt graphs as well as the 3 similarity graphs

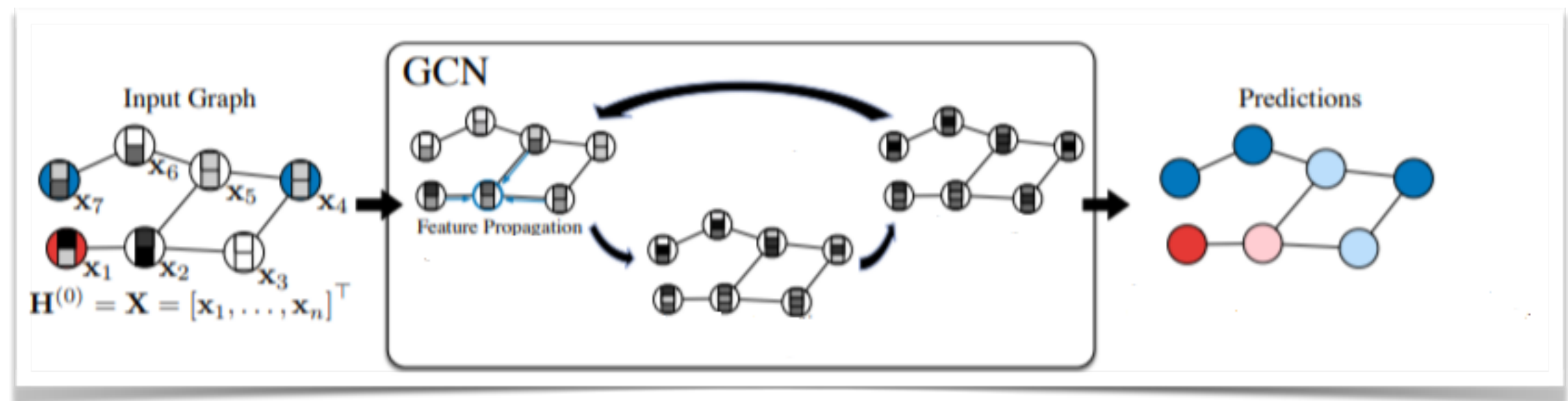


Figure 1: Structure of the convolution network

# Data Exploitation

21

---

## Third Attempt: Graph convolutional Network

Randomly splitting the data into a training and testing set

Masking 20% of the election results labels

Polynomial filter of order 3

Decay rate:  $5 \cdot 10^{-6}$

Learning rate: 0.3

# Data Exploitation

21

---

## Third Attempt: Graph convolutional Network

Randomly splitting the data into a training and testing set

Masking 20% of the election results labels

Polynomial filter of order 3

Decay rate:  $5 \cdot 10^{-6}$

Learning rate: 0.3

---

## Third Attempt: Graph convolutional Network

Randomly splitting the data into a training and testing set

Masking 20% of the election results labels

Polynomial filter of order 3

Decay rate:  $5 \cdot 10^{-6}$

Learning rate: 0.3

# Data Exploitation

21

---

## Third Attempt: Graph convolutional Network

Randomly splitting the data into a training and testing set

Masking 20% of the election results labels

Polynomial filter of order 3

Decay rate:  $5 \cdot 10^{-6}$

Learning rate: 0.3

---

## Third Attempt: Graph convolutional Network

Randomly splitting the data into a training and testing set

Masking 20% of the election results labels

Polynomial filter of order 3

Decay rate:  $5 \cdot 10^{-6}$

Learning rate: 0.3



---

## Third Attempt: Graph convolutional Network

### Results:

Graph	Accuracy Score
Return graph	91.3%
Exempt graph	92%
Total immigration sim. graph	93%
US citizen migration graph	92%
Foreigner migration graph	82%

---

## Third Attempt: Graph convolutional Network

### Results:

Graph	Accuracy Score
Return graph	91.3%
Exempt graph	92%
Total immigration sim. graph	93%
US citizen migration graph	92%
Foreigner migration graph	82%

---

## **Third Attempt: Graph convolutional Network**

No two-party clustering opportunity with all the studied features

The return and exempt graphs are scale free

GCNN performed better when large and rich data sets were used

Further improvement:

- optimise the parameters of GCNN
- add more socio-economic features on counties

# A Network Tour of Inter-County Migration in the United States

---

Q&A

# A Network Tour of Inter-County Migration in the United States

Final Project Presentation - Team 22

*Course: A Network Tour Of Data Science*

*Students: Fatima Ezzahra, Xiaoyan Zou, Paul Mansat, Anshul Toshniwal*

The logo of the École Polytechnique Fédérale de Lausanne (EPFL) is displayed in a bold, red, sans-serif font. The letters are stylized, with the 'E' and 'F' having a distinctive blocky appearance.