Dynamic Graph Machine Learning for Early Detection of Antimicrobial Resistant Outbreaks

Oskar Fraser-Krauss

Imperial College London

21st June, 2024

Introduction

– Explain disease propagation using graphs

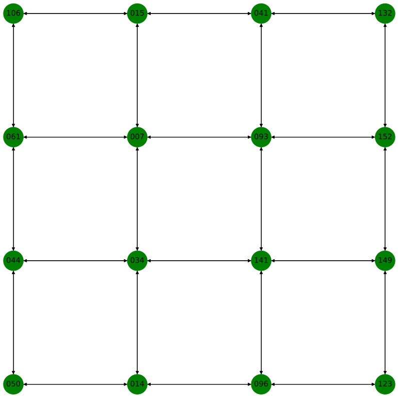
Introduction

- Explain disease propagation using graphs
- Predict spread using machine learning

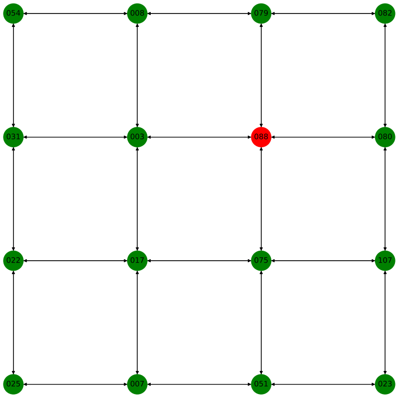
Introduction

- Explain disease propagation using graphs
- Predict spread using machine learning
- Determine intervention strategies

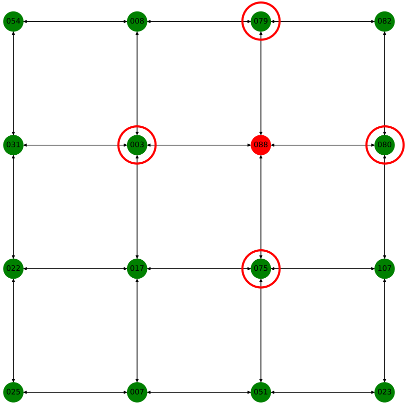
Introduction



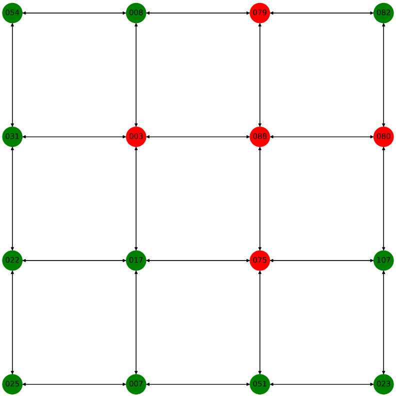
Introduction



Introduction



Introduction



Data and Graphs

Column Name	**Data Type**
patient id	`str`
admission time	`datetime`
discharge time	`datetime`
bed number	`int`
disease status	`bool`

Data and Graphs

Column Name	Data Type
patient id	`str`
admission time	`datetime`
discharge time	`datetime`
bed number	`int`
disease status	`bool`

– We build a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where the vertices (or nodes) $\mathcal{V}$ represent patients, and the edges $\mathcal{E}$ represent interactions between patients

Data and Graphs

Column Name	Data Type
patient id	`str`
admission time	`datetime`
discharge time	`datetime`
bed number	`int`
disease status	`bool`

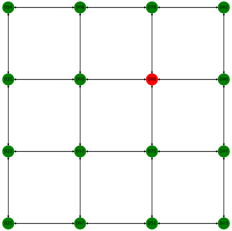
- We build a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where the vertices (or nodes) $\mathcal{V}$ represent patients, and the edges $\mathcal{E}$ represent interactions between patients
- Each node $v_i \in \mathcal{V}$ contains a set of features h_i

Dynamic Graphs

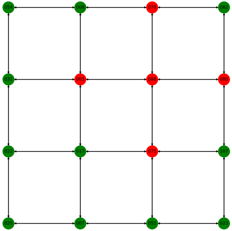
– Build daily graphs $\mathbb{G} = \{\mathcal{G}^0, \mathcal{G}^1, \ldots, \mathcal{G}^T\}$

Dynamic Graphs

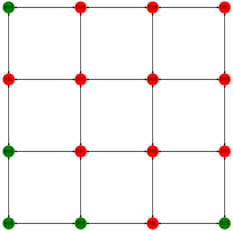
- Build daily graphs $\mathbb{G} = \{\mathcal{G}^0, \mathcal{G}^1, \ldots, \mathcal{G}^T\}$



$t = 0$ $(\mathcal{G}^0)$



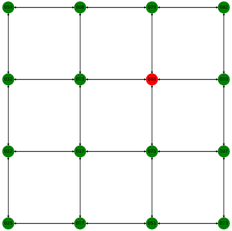
$t = 1$ $(\mathcal{G}^1)$



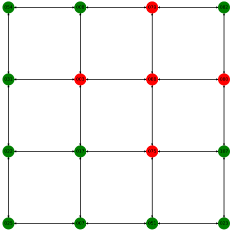
$t = 2$ $(\mathcal{G}^2)$

Dynamic Graphs

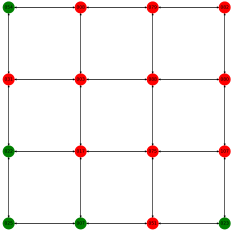
– Build daily graphs $\mathbb{G} = \{\mathcal{G}^0, \mathcal{G}^1, \ldots, \mathcal{G}^T\}$



$t = 0$ $(\mathcal{G}^0)$



$t = 1$ $(\mathcal{G}^1)$



$t = 2$ $(\mathcal{G}^2)$

– Given $\mathcal{G}^0, \mathcal{G}^1, \mathcal{G}^2$ predict $\mathcal{G}^3$

Dynamic Graphs

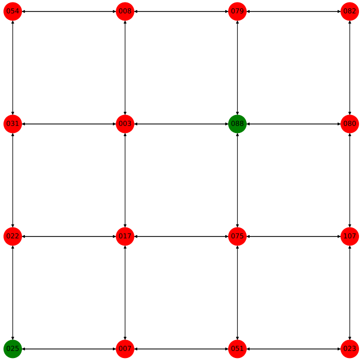


Figure: $\mathcal{G}^3$

Machine Learning

Machine Learning

- Graph neural networks (GNNs)

Machine Learning

- Graph neural networks (GNNs)
  - → Message passing

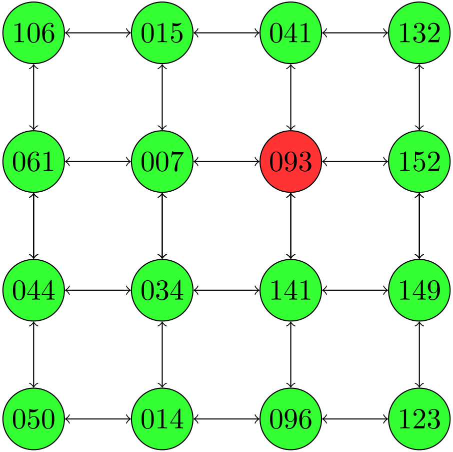
Machine Learning

- Graph neural networks (GNNs)
  - → Message passing
  - → Incorporate features of neighbouring nodes and gain information about position in the graph

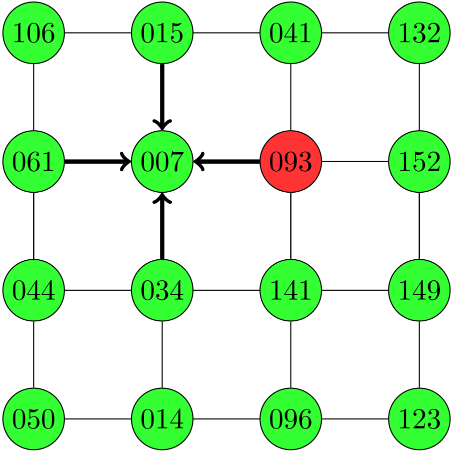
Machine Learning

- Graph neural networks (GNNs)
  - → Message passing
  - → Incorporate features of neighbouring nodes and gain information about position in the graph
- Multiply by (learnable) weight matrix and apply activation function to predict node labels

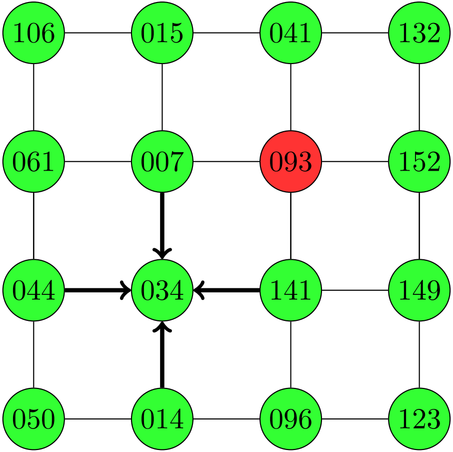
Message Passing



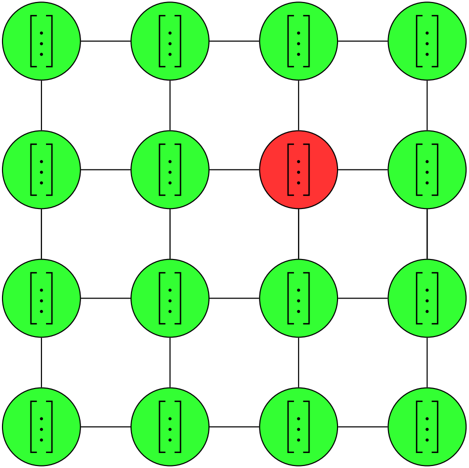
Message Passing



Message Passing



Message Passing



Message Passing

$$h_i^{(l+1)} = \left(\sum_{j \in \mathcal{N}(i)} \frac{1}{c_{ji}} h_j^{(l)} \right)$$

Message Passing

$$h_i^{(l+1)} = \left(\sum_{j \in \mathcal{N}(i)} \frac{1}{c_{ji}} h_j^{(l)} \right)$$

Message Passing

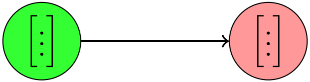
$$h_i^{(l+1)} = \left(\sum_{j \in \mathcal{N}(i)} \frac{1}{c_{ji}} h_j^{(l)} \right)$$

Message Passing

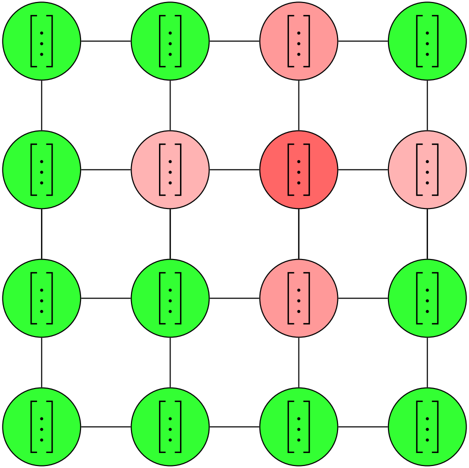
$$h_i^{(l+1)} = \left(\sum_{j \in \mathcal{N}(i)} \frac{1}{c_{ji}} h_j^{(l)} \right)$$

Message Passing

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} \frac{1}{c_{ji}} h_j^{(l)} W^{(l)}\right)$$



Message Passing



Model Training

- Simulate 100 outbreaks each 7 days long
  - → Train with 80, test with 20

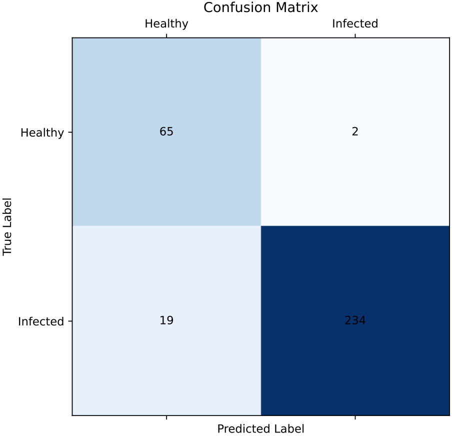
Model Training

- Simulate 100 outbreaks each 7 days long
  - → Train with 80, test with 20
- Train weights using first 5 days as features, 6th and 7th day as labels

Model Training

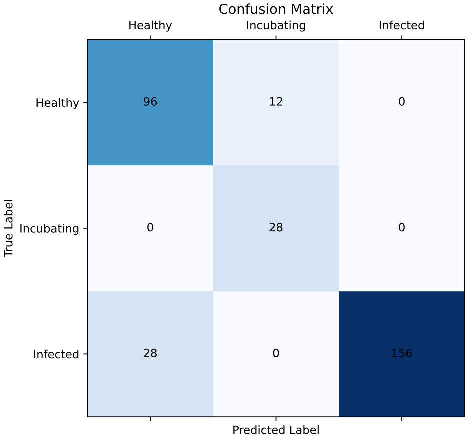
- Simulate 100 outbreaks each 7 days long
  - → Train with 80, test with 20
- Train weights using first 5 days as features, 6th and 7th day as labels
- Validate on test set

Results



Accuracy: ~ 0.95

Results



Accuracy: ~ 0.85

Summary

Summary

– Using simulated data, a basic Graph Neural Network can learn dynamics of a simple outbreak in a patient network

Summary

- Using simulated data, a basic Graph Neural Network can learn dynamics of a simple outbreak in a patient network
- High accuracy from small number of training graphs

Summary

- Using simulated data, a basic Graph Neural Network can learn dynamics of a simple outbreak in a patient network
- High accuracy from small number of training graphs
- (Roughly) captures incubation time

Next Steps

Next Steps

- Use real data!

Next Steps

- Use real data!
  - → Ideally apply to as many datasets as possible

Next Steps

- Use real data!
  - → Ideally apply to as many datasets as possible
  - → Design to accommodate real time analysis

Next Steps

- Use real data!
  - → Ideally apply to as many datasets as possible
  - → Design to accommodate real time analysis
- Model interventions

Next Steps

- Use real data!
  - → Ideally apply to as many datasets as possible
  - → Design to accommodate real time analysis
- Model interventions
- Consider more complitcated infection status

Next Steps

- Use real data!
  - → Ideally apply to as many datasets as possible
  - → Design to accommodate real time analysis
- Model interventions
- Consider more complitcated infection status
  - → Ideally this would be a label with the type of infection or detailed genomic data

Next Steps

- Use real data!
  - → Ideally apply to as many datasets as possible
  - → Design to accommodate real time analysis
- Model interventions
- Consider more complitcated infection status
  - → Ideally this would be a label with the type of infection or detailed genomic data
- Include more detailed patient information

Next Steps

- Use real data!
  - → Ideally apply to as many datasets as possible
  - → Design to accommodate real time analysis
- Model interventions
- Consider more complitcated infection status
  - → Ideally this would be a label with the type of infection or detailed genomic data
- Include more detailed patient information
  - → e.g. age, comorbidities

Next Steps

- Use real data!
  - → Ideally apply to as many datasets as possible
  - → Design to accommodate real time analysis
- Model interventions
- Consider more complitcated infection status
  - → Ideally this would be a label with the type of infection or detailed genomic data
- Include more detailed patient information
  - → e.g. age, comorbidities
- Extend beyond a single ward

Further Thoughts

Further Thoughts

– Incorporate different layouts

Further Thoughts

- Incorporate different layouts
- Modelled without in-depth understanding of how wards are run

Questions?

– Any questions?