Predicting Missing Links in Infection Networks: Accelerate contact tracing investigations using network theory.

Pavlos Alexandros Dimitriou[1,2], V. Silvestros[3], C.Pitris[1,2], P. Kolios[1,4]

[1]University of Cyprus, KIOS Research and Innovation Center of Excellence, Nicosia, Cyprus.

[2]University of Cyprus, Dept. of Electrical & Computer Engineering, Nicosia, Cyprus.

[3]Unit for Surveillance and Control of Communicable Diseases, Ministry of Health, Nicosia, Cyprus.

[4]University of Cyprus, Dept. of Computer Science, Nicosia, Cyprus.

24-27 September, World Congress of Epidemiology

University of Cyprus

WORLD CONGRESS OF EPIDEMIOLOGY 2024

Building networks from contact tracing data

Data store in excel tables

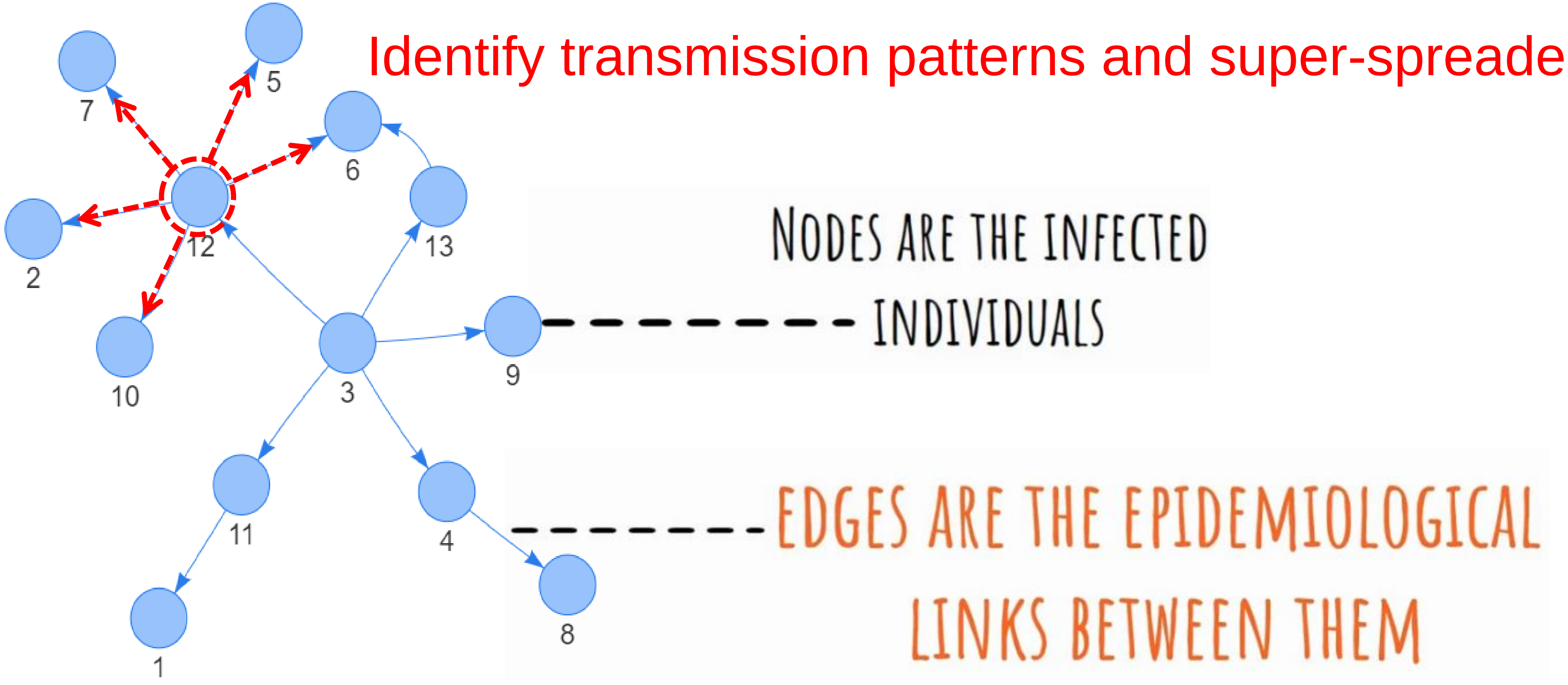
Case Id	Contact case Id	Gender	District	Age group	NACE	Infection date
1	11	M	Pafos	20-29	P85.4	12/1/2021
2	12	M	Pafos	3-5	U6	12/1/2021
3	nan	M	Pafos	30-39	P85.4	12/1/2021
4	3	F	Pafos	40-49	P85.4	12/1/2021
5	12	M	Pafos	3-5	U6	12/1/2021
6	12,13	F	Pafos	0-2	U6	12/1/2021
7	12	F	Pafos	30-39	nan	12/1/2021
8	4	F	Pafos	40-49	nan	12/1/2021
9	3	F	Pafos	60-69	R65	12/1/2021
10	12	M	Pafos	20-29	T97	12/1/2021
11	3	F	Pafos	50-59	P85.4	12/1/2021
12	3	F	Pafos	60-69	R65	12/1/2021
13	3	M	Pafos	70-79	R65	12/1/2021

Challenging to analyse

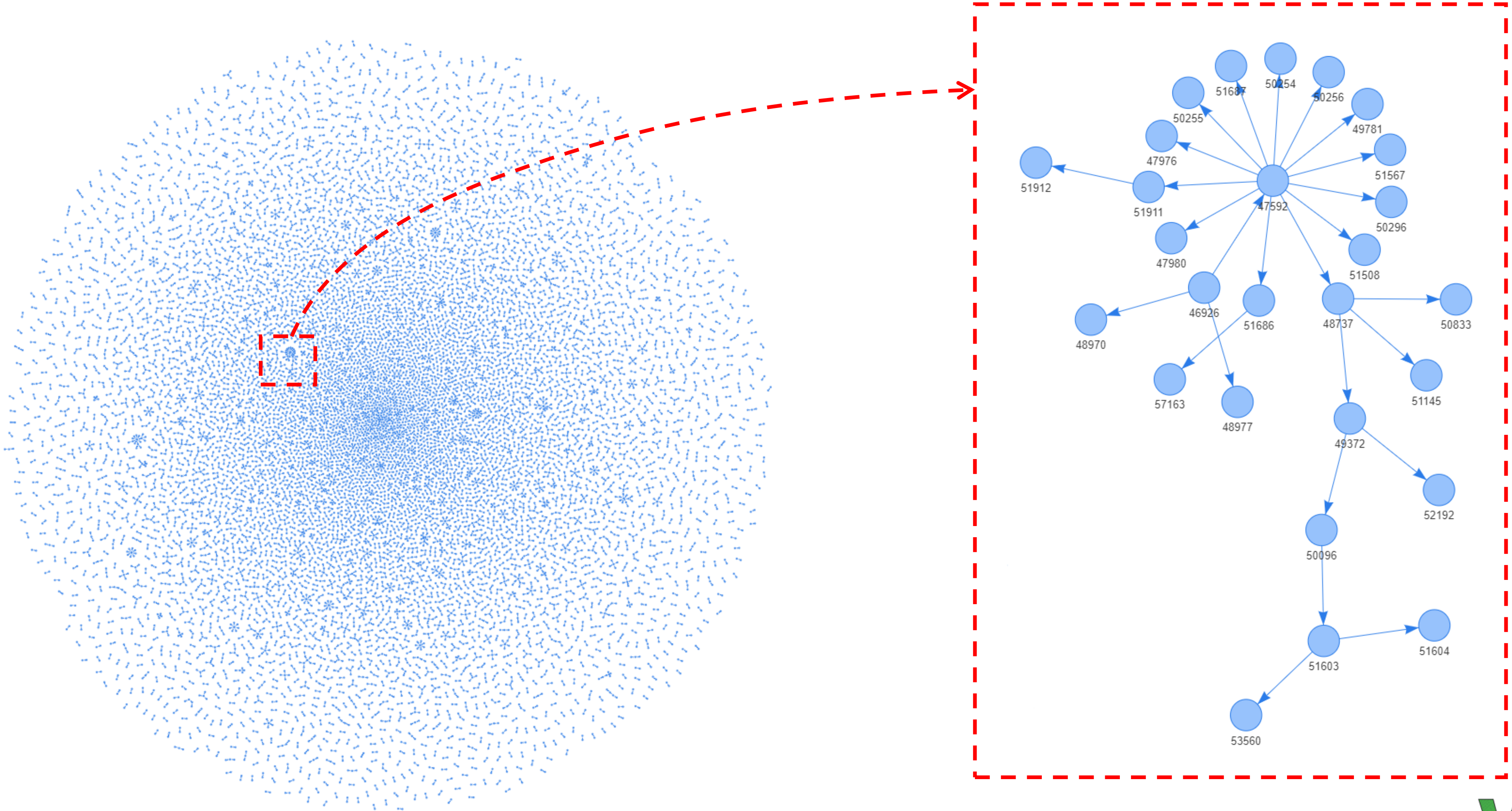
Building networks from contact tracing data

Infection network:

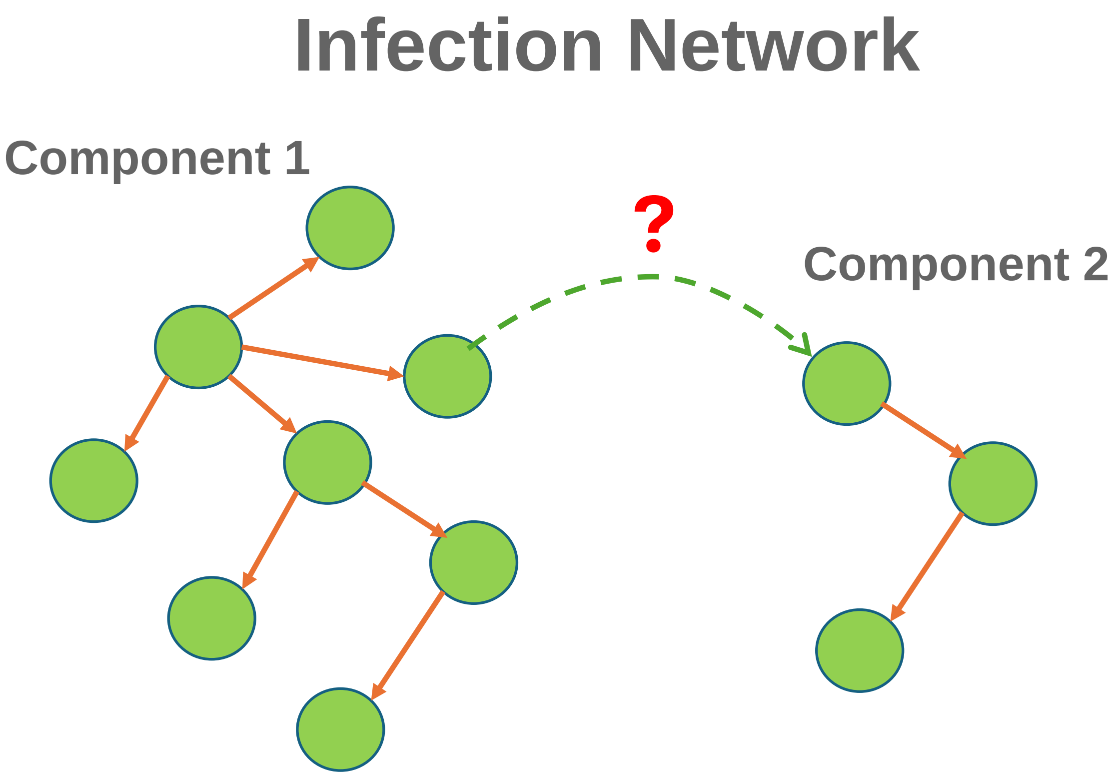
Identify transmission patterns and super-spreaders



Infection networks are sparse



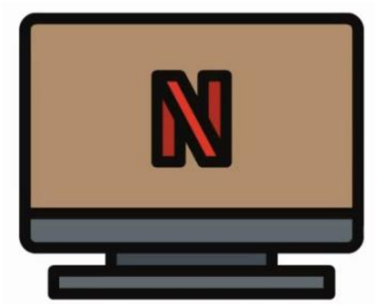
Can we predict missing links?

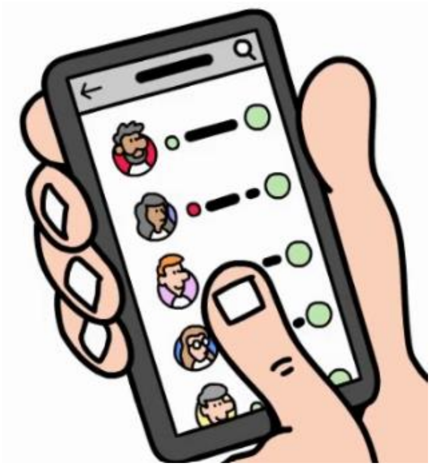


Link prediction

Link Prediction and Its Application:

- Widely used in different domains
  - Recommender system
    - friendships in social networks
    - e-commerce websites





Feature vector (x)

$$x = [f_{0,0}, \dots f_{i,j}, \dots, f_{n,n}]$$

Class label (y)

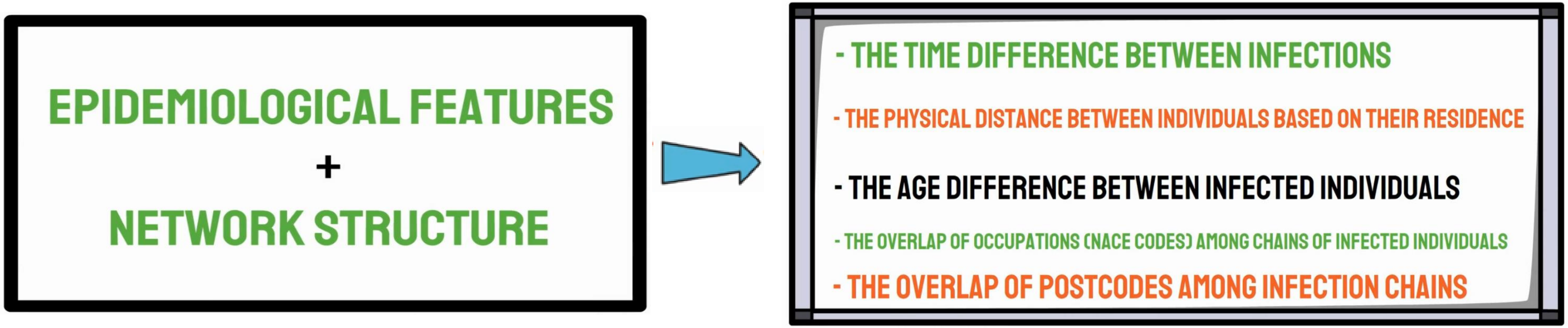
$$y = [y_{0,0}, \dots y_{i,j}, \dots, y_{n,n}]$$

Creating Feature Vectors:

1. Node2vec:

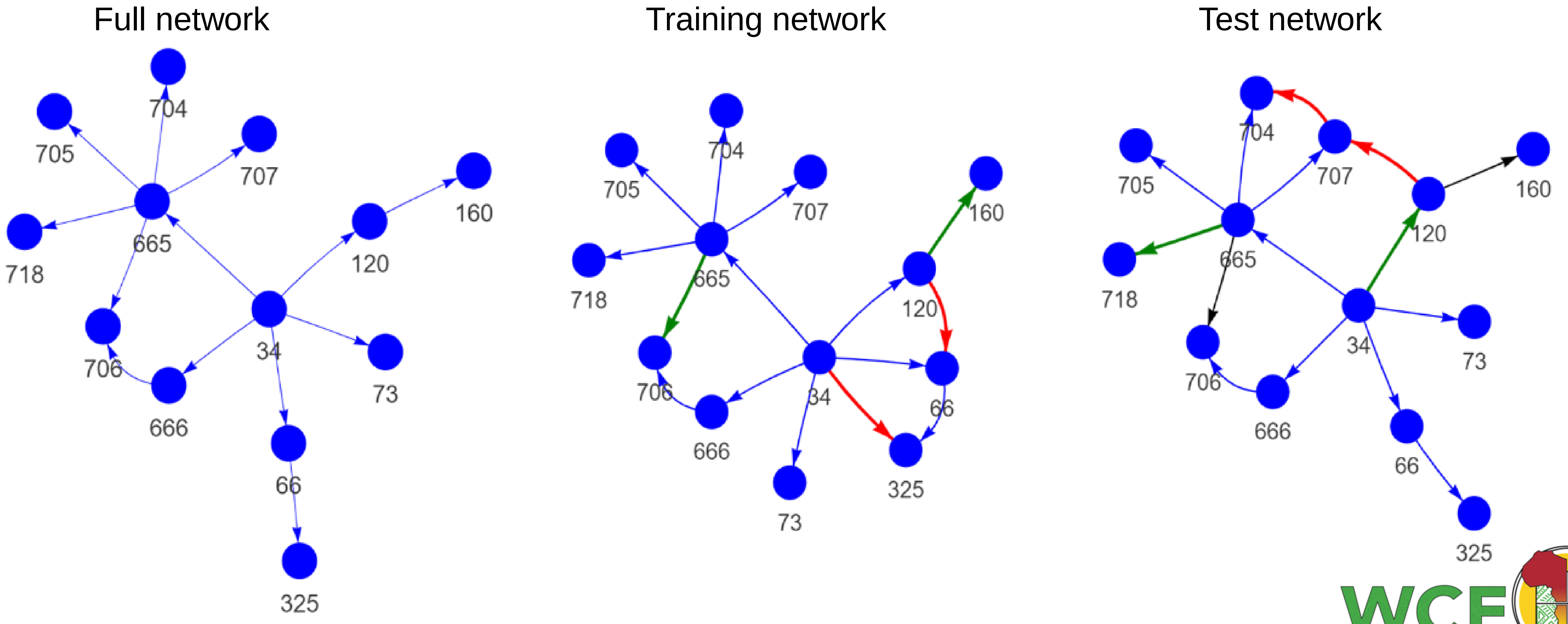
- Features are calculated solely on network characteristics

2. Shallow embeddings with handcrafted features:



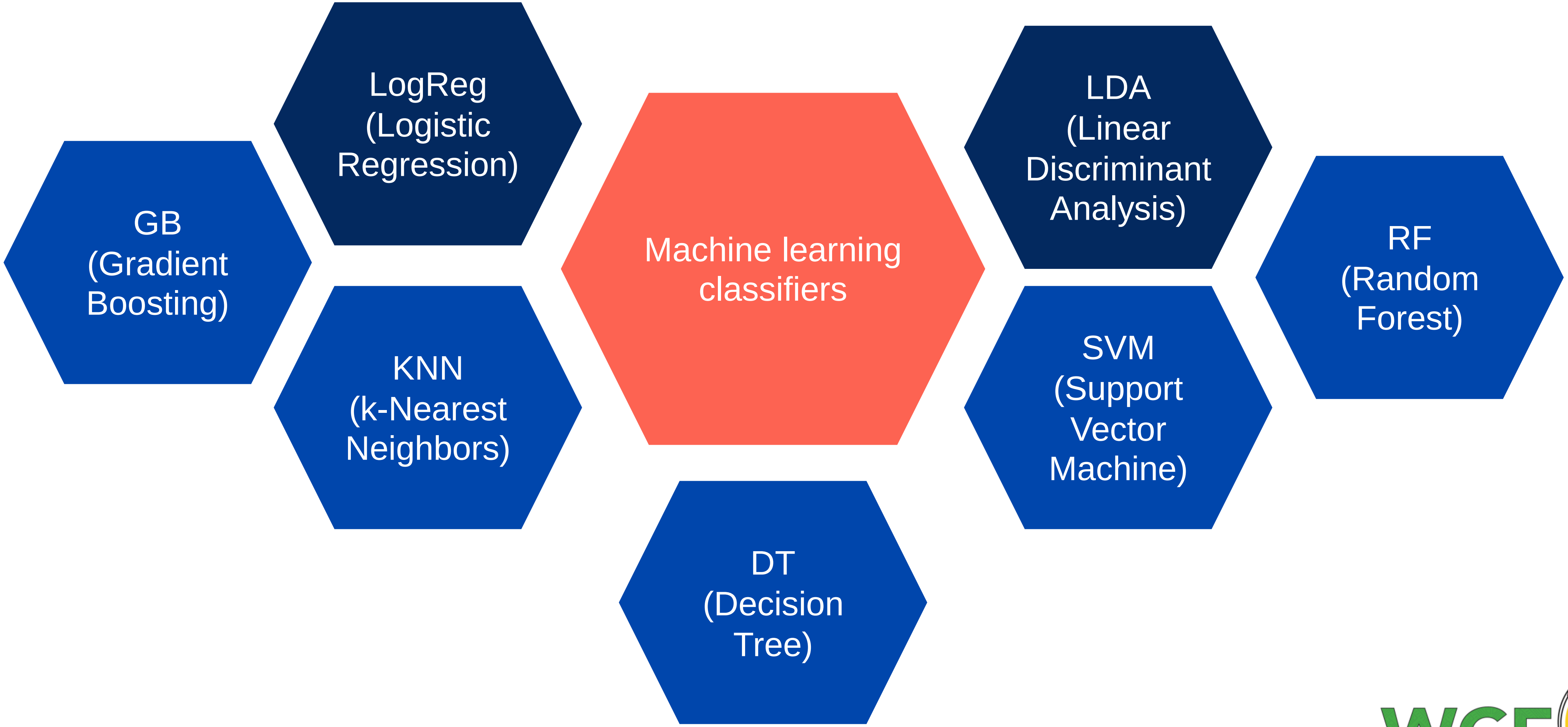
Splitting the Dataset

- Sampling positive and negative edges to create training and test set



Alexandros Dimitriou, Cyprus

Machine Learning: Classification algorithms



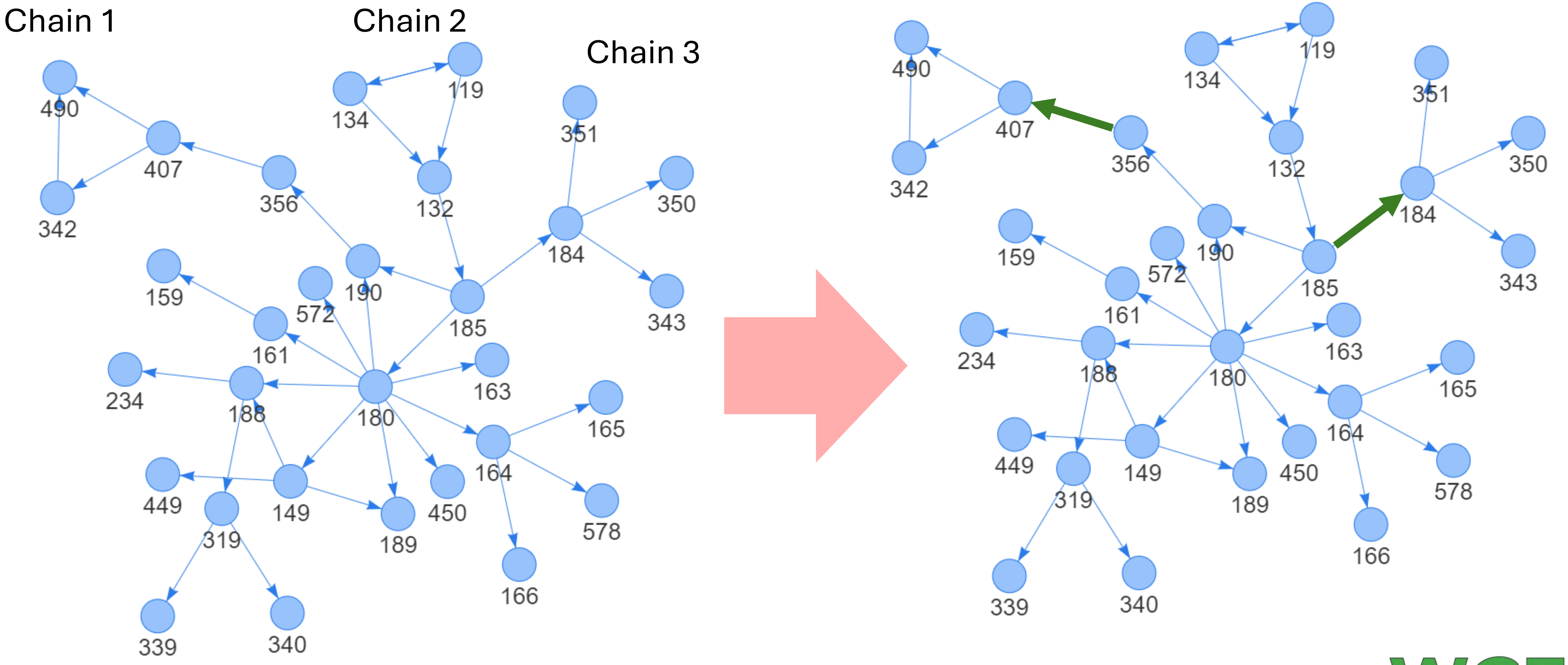
Performance Metrics: Node2vec Link prediction

	Wave 1		Wave 2		Wave 3		Wave 4	
	AUC	F1-Score	AUC	F1-Score	AUC	F1-Score	AUC	F1-Score
RF	0.54	0.49	0.54	0.46	0.57	0.53	0.67	0.65
LDA	0.50	0.51	0.48	0.47	0.55	0.57	0.65	0.71
SVM	0.54	0.53	0.50	0.51	0.61	0.64	**0.70**	**0.75**
LOGREG	0.53	0.52	0.49	0.49	0.58	0.60	0.68	0.75
KNN	0.56	0.61	0.49	0.50	0.58	0.65	0.64	0.62
NB	0.52	0.42	0.51	0.47	0.58	0.55	0.62	0.71
GB	0.57	0.57	0.52	0.52	0.58	0.61	0.69	0.71
DT	0.53	0.52	0.50	0.51	0.54	0.54	0.62	0.62

Performance Metrics:
Shallow embeddings with handcrafted features

	Wave 1		Wave 2		Wave 3		Wave 4	
	AUC	F1-Score	AUC	F1-Score	AUC	F1-Score	AUC	F1-Score
RF	0.87	0.85	0.82	0.82	0.93	0.93	0.94	0.94
LDA	0.85	0.86	**0.82**	**0.83**	0.90	0.90	0.91	0.91
SVM	0.56	0.57	0.55	0.61	0.54	0.59	0.54	0.53
LOGREG	0.54	0.17	0.53	0.24	0.53	0.10	0.53	0.11
KNN	0.78	0.75	0.74	0.75	0.67	0.67	0.73	0.73
NB	0.56	0.57	0.55	0.58	0.54	0.61	0.53	0.63
GB	**0.90**	**0.88**	0.82	0.82	**0.94**	**0.94**	**0.95**	**0.95**
DT	0.81	0.85	0.79	0.80	0.91	0.90	0.92	0.92

Link prediction: Complete transmissions chains



Alexandros Dimitriou, Cyprus

Grant agreement No. 739551 (KIOS CoE).
CIPHIS project (Cyprus Innovative Public Health ICT System), C1.1l2, of the NextGenerationEU-Recovery and Resilience Plan