Entity Linkage in Evolving Temporal Graphs

Data Intensive Systems - 2024

Utrecht University

Group #15

Instructor: Prof. Yannis Velegrakis

Alex Francis

Bar Melinarskiy

Sofia Bianchini

DATASETS

01 Initial small dataset – testing spatial & temporal similarity

Attribute	Size
Samples	15
Communities	3
Groups	2
Community range	3-5

02 Increased sample size with small communities

Attribute	Size
Samples	5000
Communities	200
Groups	20
Community range	3-5

03 Increased community size range

Attribute	Size
Samples	5000
Communities	30
Groups	20
Community range	3-500

04 Big dataset

Attribute	Size
Samples	50000
Communities	500
Groups	200
Community range	3-5000

TECHNOLOGIES USED

SOLUTION GENERAL STEPS

Step 1

Generate datasets

Step 2

Create preprocessing pipeline

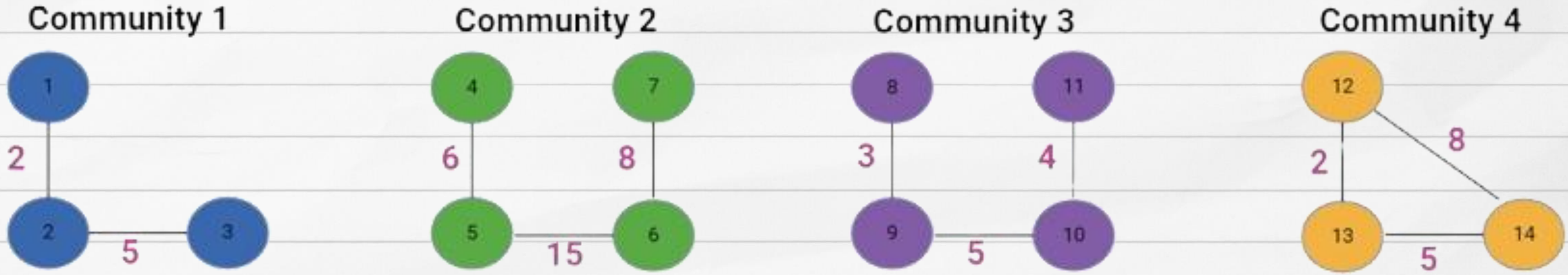
Step 3

Extract similar groups

Step 4

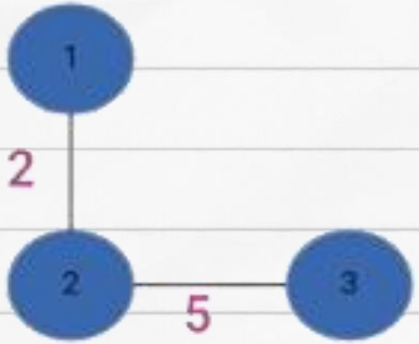
Performance evaluation

ADJACENCY MATRIX – TOY EXAMPLE



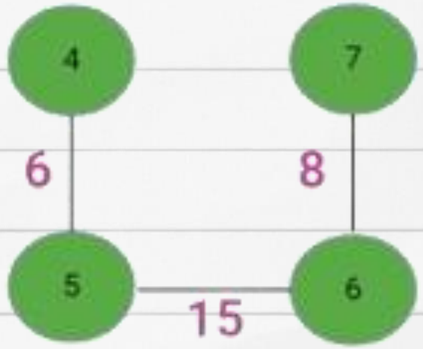
ADJACENCY MATRIX – TOY EXAMPLE

Community 1



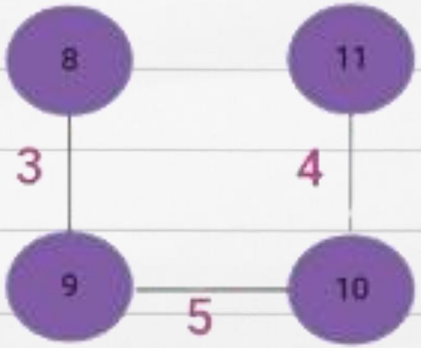
	1	2	3
1	0	2	0
2	2	0	5
3	0	5	0

Community 2



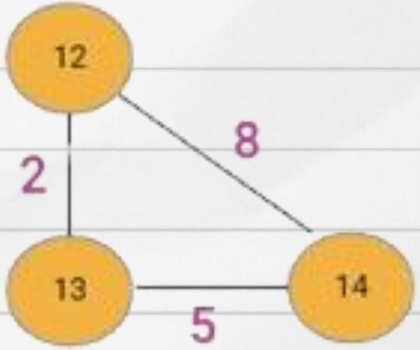
	4	5	6	7
4	0	6	0	0
5	6	0	15	0
6	0	15	0	8
7	0	0	8	0

Community 3



	8	9	10	11
8	0	3	0	0
9	3	0	5	0
10	0	5	0	4
11	0	0	4	0

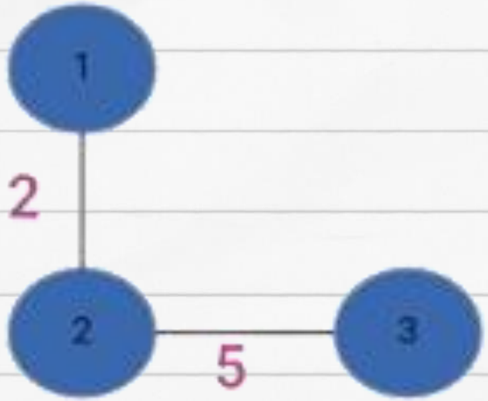
Community 4



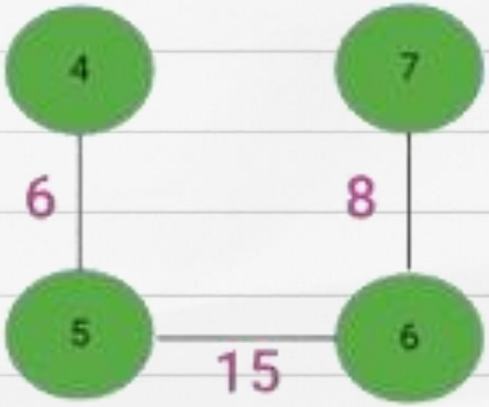
	12	13	14
12	0	2	8
13	2	0	5
14	8	5	0

ADJACENCY MATRIX – TOY EXAMPLE

Community 1



Community 2

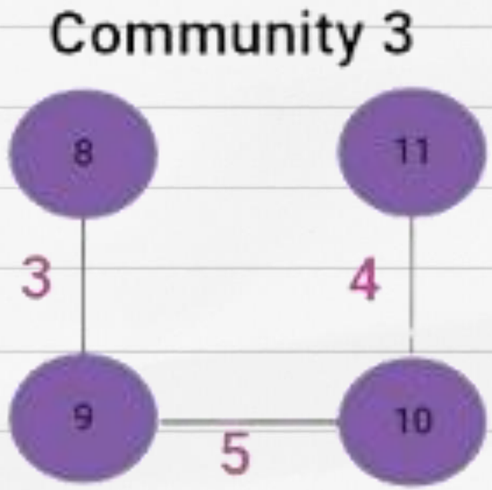


Communities Pair	Similarity Score
1 + 2	0.90

	1	2	3
1	0	2	0
2	2	0	5
3	0	5	0

	4	5	6	7
4	0	6	0	0
5	6	0	15	0
6	0	15	0	8
7	0	0	8	0

ADJACENCY MATRIX – TOY EXAMPLE



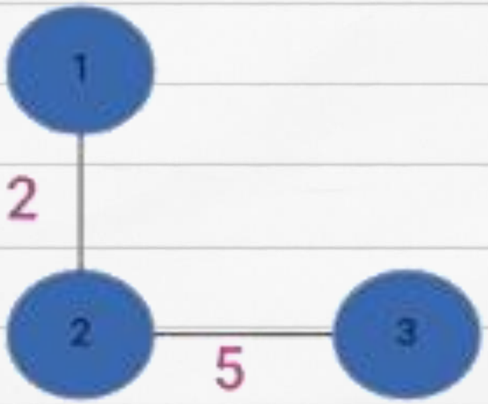
Communities Pair	Similarity Score
2 + 3	0.98

	4	5	6	7
4	0	6	0	0
5	6	0	15	0
6	0	15	0	8
7	0	0	8	0

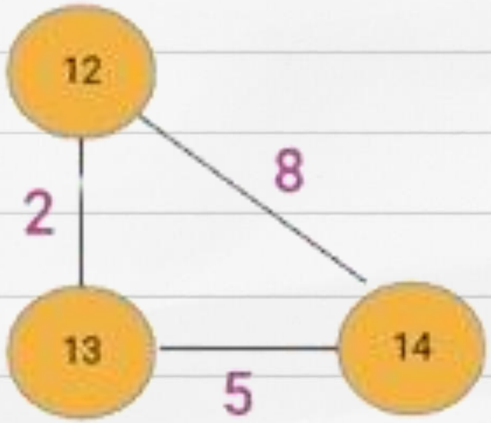
	8	9	10	11
8	0	3	0	0
9	3	0	5	0
10	0	5	0	4
11	0	0	4	0

ADJACENCY MATRIX – TOY EXAMPLE

Community 1



Community 4



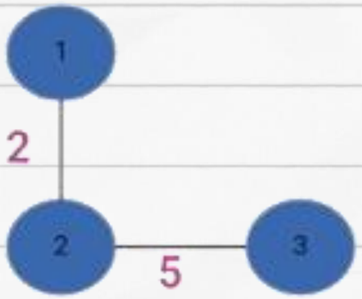
Communities Pair	Similarity Score
1 + 4	0.50

	1	2	3
1	0	2	0
2	2	0	5
3	0	5	0

	12	13	14
12	0	2	8
13	2	0	5
14	8	5	0

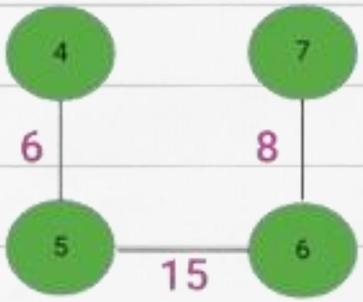
ADJACENCY MATRIX – TOY EXAMPLE

Community 1



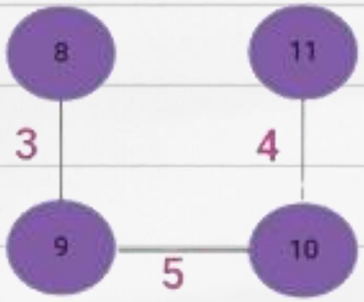
	1	2	3
1	0	2	0
2	2	0	5
3	0	5	0

Community 2



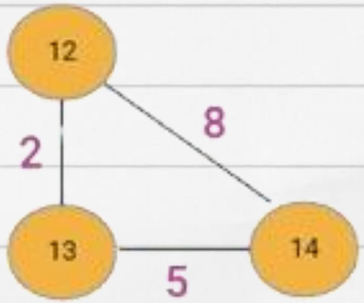
	4	5	6	7
4	0	6	0	0
5	6	0	15	0
6	0	15	0	8
7	0	0	8	0

Community 3



	8	9	10	11
8	0	3	0	0
9	3	0	5	0
10	0	5	0	4
11	0	0	4	0

Community 4



	12	13	14
12	0	2	8
13	2	0	5
14	8	5	0

Communities Pair	Similarity Score
1 + 2	0.90
1 + 3	0.81
1 + 4	0.56
2 + 3	0.98
2 + 4	0.50
3 + 4	0.45

With 0.8 cutoff threshold

THANK YOU FOR LISTENING!