

India Covid19 TDA using Mapper

Likhith Marrapu
Vijay Kumar Boda
Yash Bagdi

Introduction

- TDA
- Shapes in TDA
- When to use TDA
- Need for TDA



Mapper algorithm

- **Project:** a dataset to an Euclidean Space.
- **Cover:** this projection with overlapping intervals
- **Cluster:** the points inside these intervals i.e. covers
 - Clustering done on original data features
 - Clusters becomes nodes in graph
- **Edge** is drawn: between to nodes if their respective clusters have common points
- For extra visualization:
 - Change size, colour, shape of nodes/edges according to required functionalities
 - Provide descriptive statistics on nodes and graph



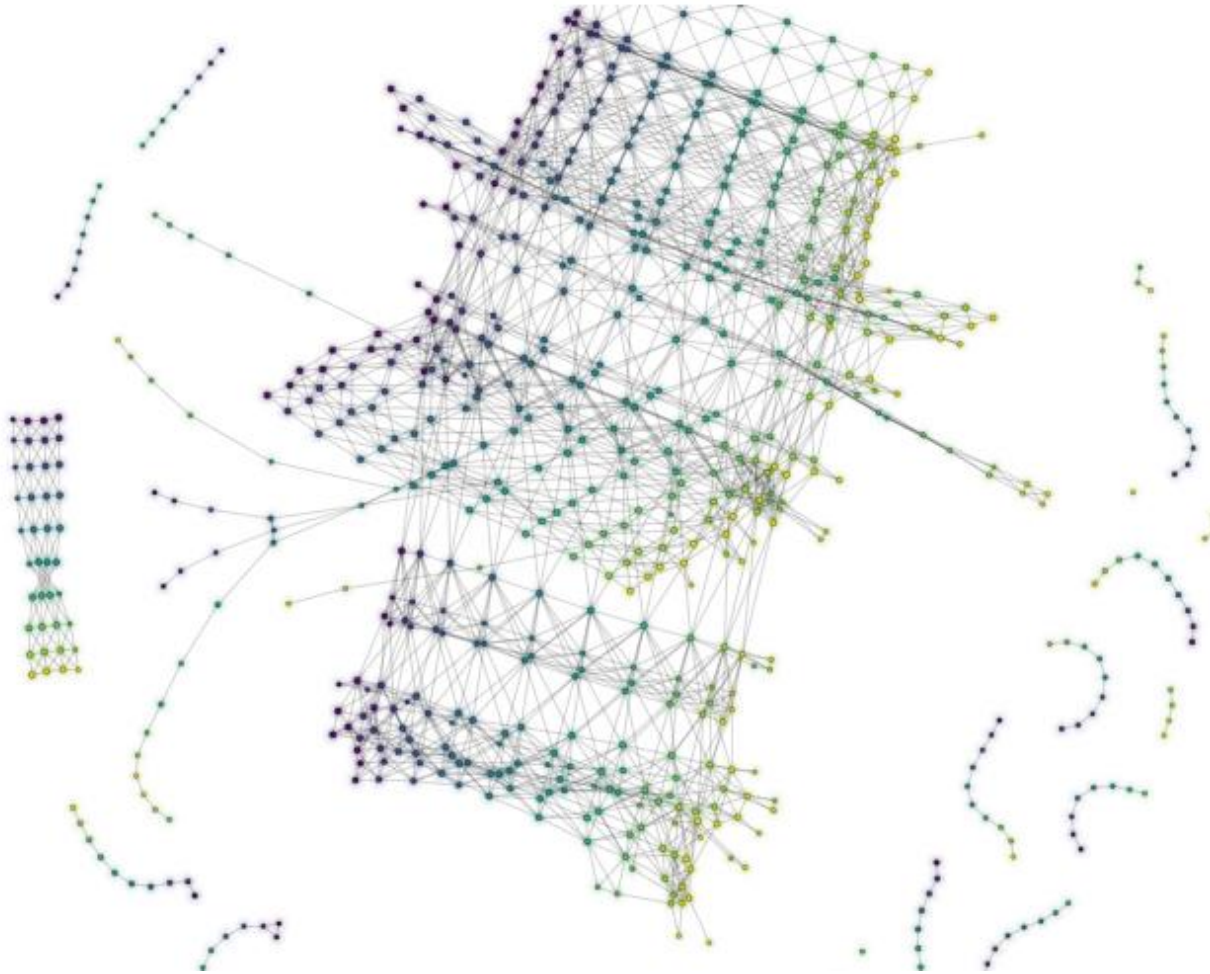
Dataset

- COVID-19 ,INDIA (<https://api.covid19india.org/>).
- District wise timeseries of confirmed covid cases in INDIA.
- Timeseries range: 1 January 2021 to 31 May 2021.
- Each datapoints contains:
 - Date
 - District and State Names
 - Confirmed , Recovered and deaths

Pre-Processing Covid-19 Data

- Every point " p "
- $p = (\text{Latitude}, \text{Longitude}, \# \text{ of Cumulative Confirmed Cases}, \text{Day})$
- Normalization:
 - Feature wise for the duration chosen.
- Applying Mapper:
 - Projection: identity map, $f: X \rightarrow \mathbb{R}^4$
 - Cover: Default KeplerMapper Cover with $n = 10$, $\delta = 0.08$
 - Clustering Algorithm: DBSCAN with $eps = 0.02$, $min_samples = 8$
 - Best " eps " from parameter estimation technique.

Mapper Summary



Mapper Summary

PROJECTION [0, 1, 2, 3]
N_CUBES 10
PERC_OVERLAP 0.08
CLUSTERER DBSCAN(eps=0.02,
min_samples=8)
SCALER MinMaxScaler()
NODES 756
EDGES 2896
TOTAL SAMPLES 128682
UNIQUE SAMPLES 101319

NODE DISTRIBUTION



Cluster Details

Cluster Details (node id:
cube373_cluster0)

MEMBER DISTRIBUTION

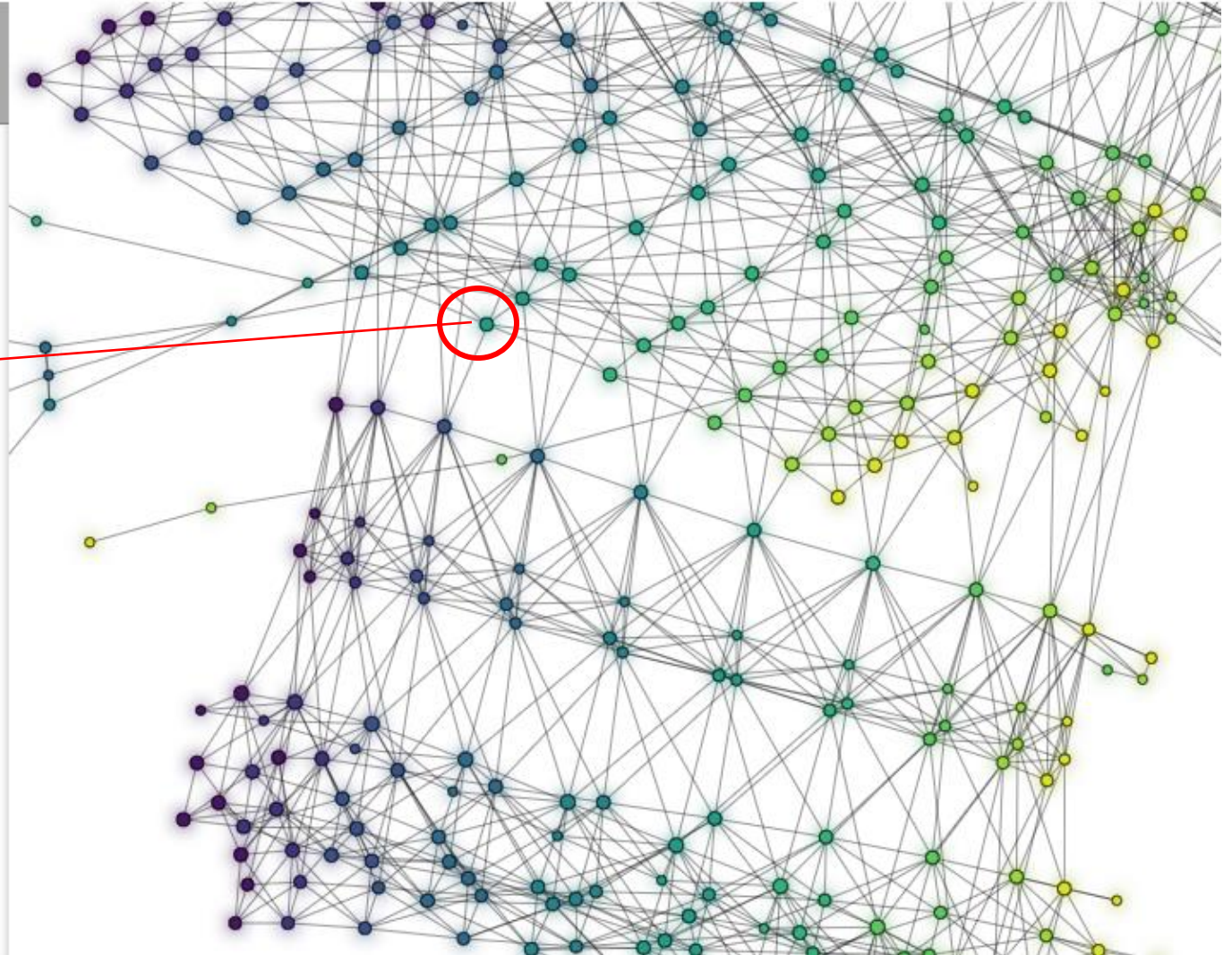
0.0% 0.0% 0.0% 0.0% 5.9% 88.2% 5.9% 0.0% 0.0% 0.0%

SIZE

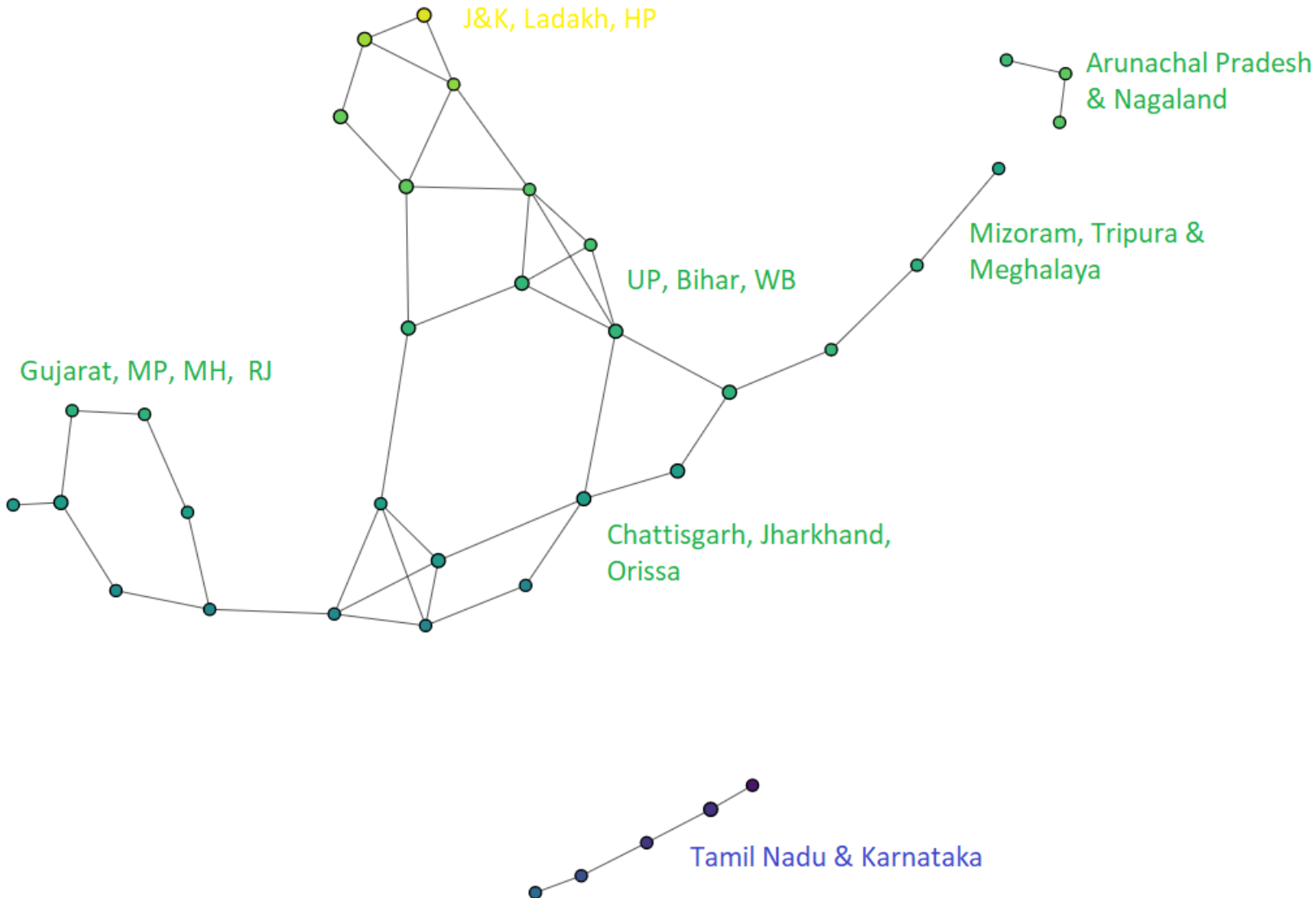
187

MEMBERS

Botad, Gujarat, 2021-03-19, 1066;
Devbhumi Dwarka, Gujarat, 2021-03-
19, 1178; Jamnagar, Gujarat, 2021-03-
19, 10943; Junagadh, Gujarat, 2021-03-
19, 5607; Kutch, Gujarat, 2021-03-



Effect of Geography and Number of Cases



Mapper Summary

PROJECTION $[0, 1, 2]$

N_CUBES 10

PERC_OVERLAP 0.08

```
CLUSTERER DBSCAN(eps=0.02,  
min_samples=8)
```

SCALER MinMaxScaler()

NODES 36

EDGES 47

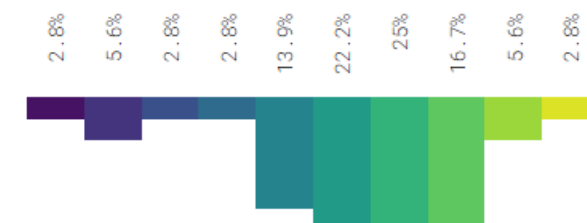
TOTAL SAMPLES 643

UNIQUE SAMPLES 561

COLOR FUNCTIONS Lat, Long, Confirmed

NODE COLOR FUNCTION mean

NODE DISTRIBUTION



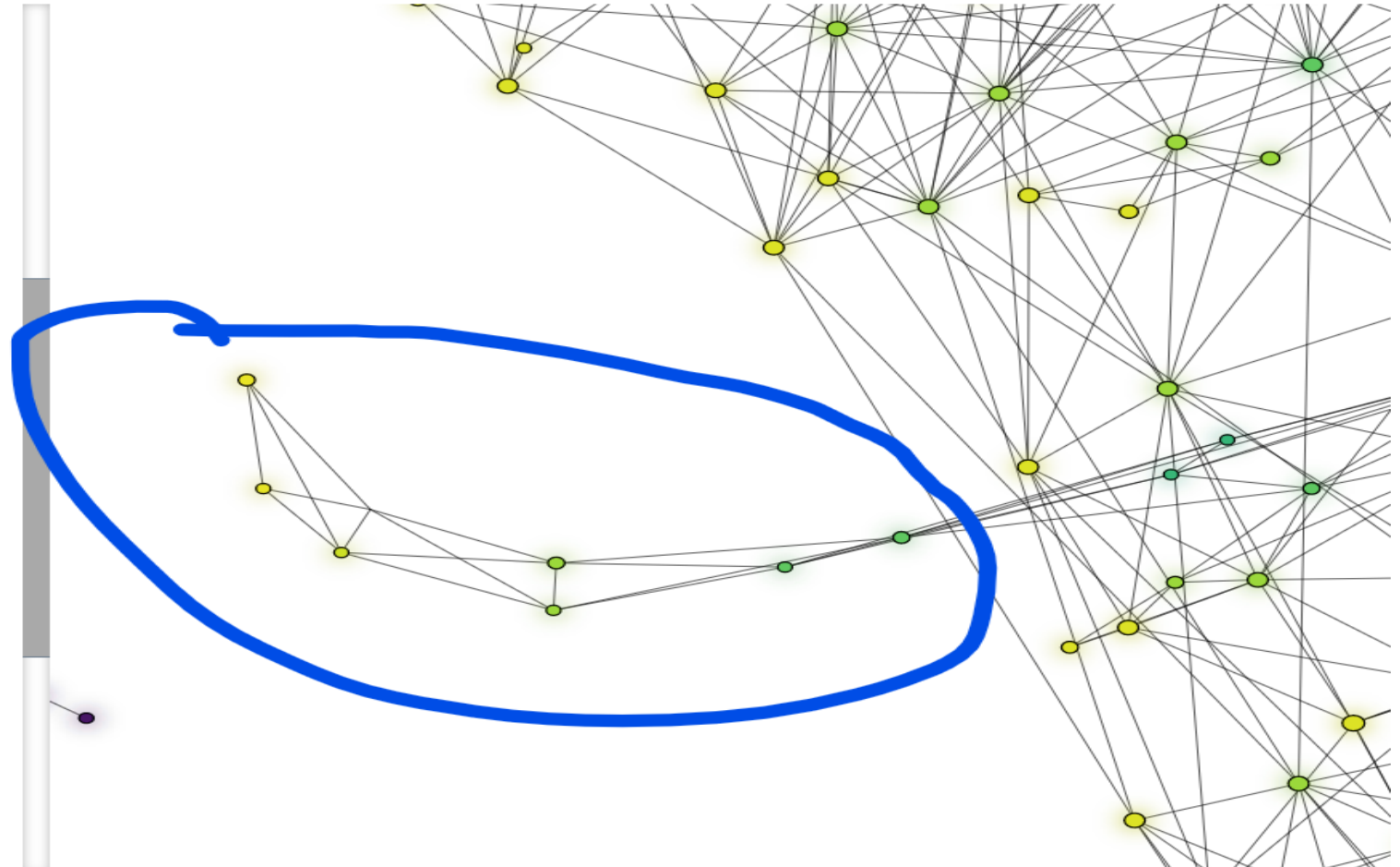
Flares: Outbreak Situation

COVID DATA

22

MEMBERS

Kolkata, West Bengal, 2021-05-19, 265376
North 24 Parganas, West Bengal, 2021-05-19, 254630
Kolkata, West Bengal, 2021-05-20, 268837
North 24 Parganas, West Bengal, 2021-05-20, 258748
Kolkata, West Bengal, 2021-05-21, 272397
North 24 Parganas, West Bengal, 2021-05-21, 262988
Kolkata, West Bengal, 2021-05-22, 275677
North 24 Parganas, West Bengal, 2021-05-22, 266864
Kolkata, West Bengal, 2021-05-23, 278733
North 24 Parganas, West Bengal, 2021-05-23, 270635
Kolkata, West Bengal, 2021-05-24, 281854
North 24 Parganas, West Bengal, 2021-05-24, 274428
Kolkata, West Bengal, 2021-05-25, 284833
North 24 Parganas, West Bengal, 2021-05-25, 277880
Kolkata, West Bengal, 2021-05-



COVID DATA

19

MEMBERS

Mumbai, Maharashtra, 2021-05-19, 691352
Mumbai, Maharashtra, 2021-05-20, 692785
Mumbai, Maharashtra, 2021-05-21, 694200
Mumbai, Maharashtra, 2021-05-22, 695483
Mumbai, Maharashtra, 2021-05-23, 696910
Mumbai, Maharashtra, 2021-05-24, 697959
Mumbai, Maharashtra, 2021-05-25, 698988
Mumbai, Maharashtra, 2021-05-26, 700340
Mumbai, Maharashtra, 2021-05-27, 701598
Mumbai, Maharashtra, 2021-05-28, 702522
Mumbai, Maharashtra, 2021-05-29, 703560
Mumbai, Maharashtra, 2021-05-30, 704622
Mumbai, Maharashtra, 2021-05-31, 705288
Mumbai, Maharashtra, 2021-06-01, 706118
Mumbai, Maharashtra, 2021-06-02, 707041



Segmented Flares: #of cases or Geography

COVID DATA

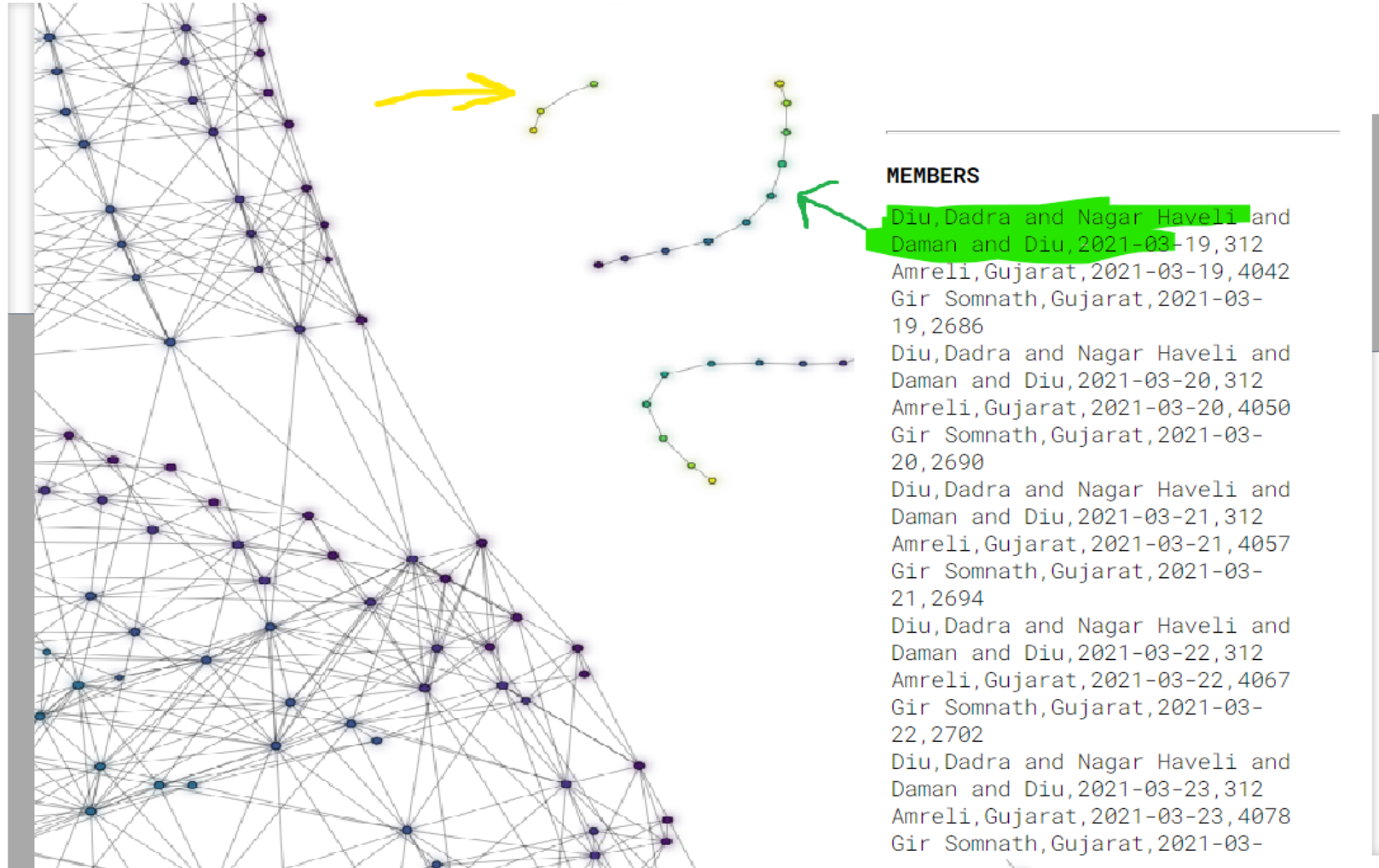
CLUSTER STATISTICS

SIZE

9

MEMBERS

Bengaluru Urban, Karnataka, 2021-05-12, 999805
Bengaluru Urban, Karnataka, 2021-05-13, 1014996
Bengaluru Urban, Karnataka, 2021-05-14, 1029312
Bengaluru Urban, Karnataka, 2021-05-15, 1042714
Bengaluru Urban, Karnataka, 2021-05-16, 1051058
Bengaluru Urban, Karnataka, 2021-05-17, 1064396
Bengaluru Urban, Karnataka, 2021-05-18, 1073072
Bengaluru Urban, Karnataka, 2021-05-19, 1084844
Bengaluru Urban, Karnataka, 2021-05-20, 1094253



Flare Complex

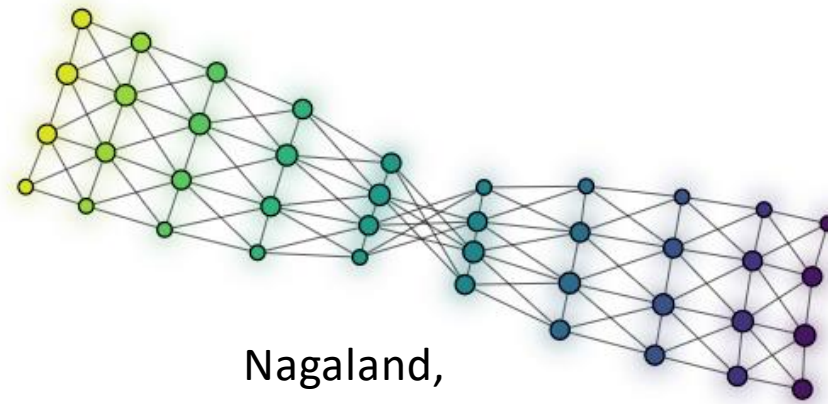
COVID DATA

SIZE

130

MEMBERS

Unknown, Manipur, 2021-04-06, 29435;
Kiphire, Nagaland, 2021-04-06, 41;
Kohima, Nagaland, 2021-04-06, 4029;
Longleng, Nagaland, 2021-04-06, 19;
Mokokchung, Nagaland, 2021-04-06, 195;
Mon, Nagaland, 2021-04-06, 605;
Phek, Nagaland, 2021-04-06, 47;
Tuensang, Nagaland, 2021-04-06, 250;
Wokha, Nagaland, 2021-04-06, 39;
Zunheboto, Nagaland, 2021-04-06, 142;
Unknown, Manipur, 2021-04-07, 29447;
Kiphire, Nagaland, 2021-04-07, 41;
Kohima, Nagaland, 2021-04-07, 4035;
Longleng, Nagaland, 2021-04-07, 19;
Mokokchung, Nagaland, 2021-04-07, 195;
Mon, Nagaland, 2021-04-07, 605;
Phek, Nagaland, 2021-04-07, 47;
Tuensang, Nagaland, 2021-04-07, 250;
Wokha, Nagaland, 2021-04-07, 39;
Zunheboto, Nagaland, 2021-04-07, 142;
Unknown, Manipur, 2021-04-08, 29463;
Kiphire, Nagaland, 2021-04-08, 41;



Nagaland,
Manipur,
Arunachal Pradesh

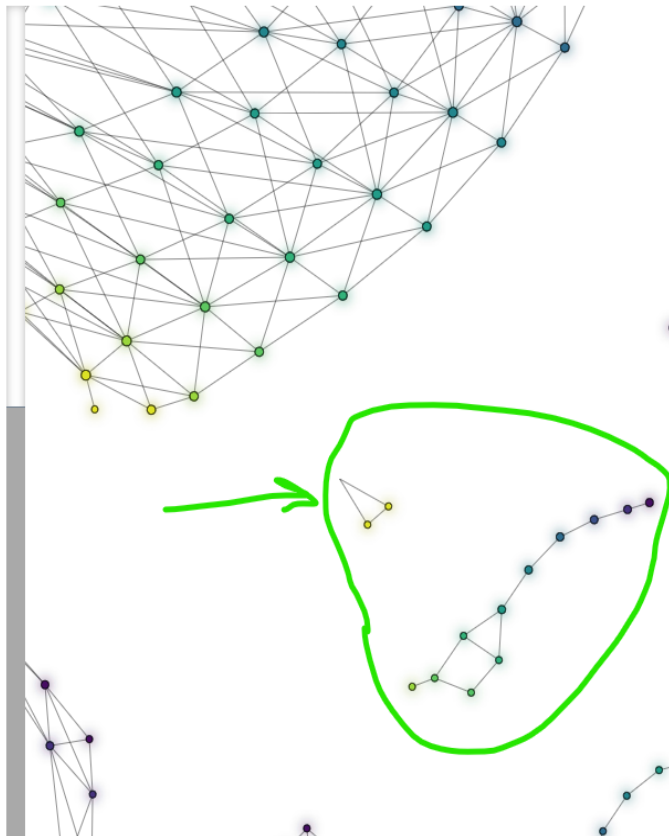
Reconnected Flares

Jan to Apr

COVID DATA

MEMBERS

Mumbai, Maharashtra, 2021-04-06, 472600
Mumbai, Maharashtra, 2021-04-07, 483042
Mumbai, Maharashtra, 2021-04-08, 491980
Mumbai, Maharashtra, 2021-04-09, 501182
Mumbai, Maharashtra, 2021-04-10, 510512
Mumbai, Maharashtra, 2021-04-11, 520498
Mumbai, Maharashtra, 2021-04-12, 527391
Mumbai, Maharashtra, 2021-04-13, 535264
Mumbai, Maharashtra, 2021-04-14, 545195
Mumbai, Maharashtra, 2021-04-15, 553404
Mumbai, Maharashtra, 2021-04-16, 562207
Mumbai, Maharashtra, 2021-04-17, 571018
Mumbai, Maharashtra, 2021-04-18, 579486



COVID DATA

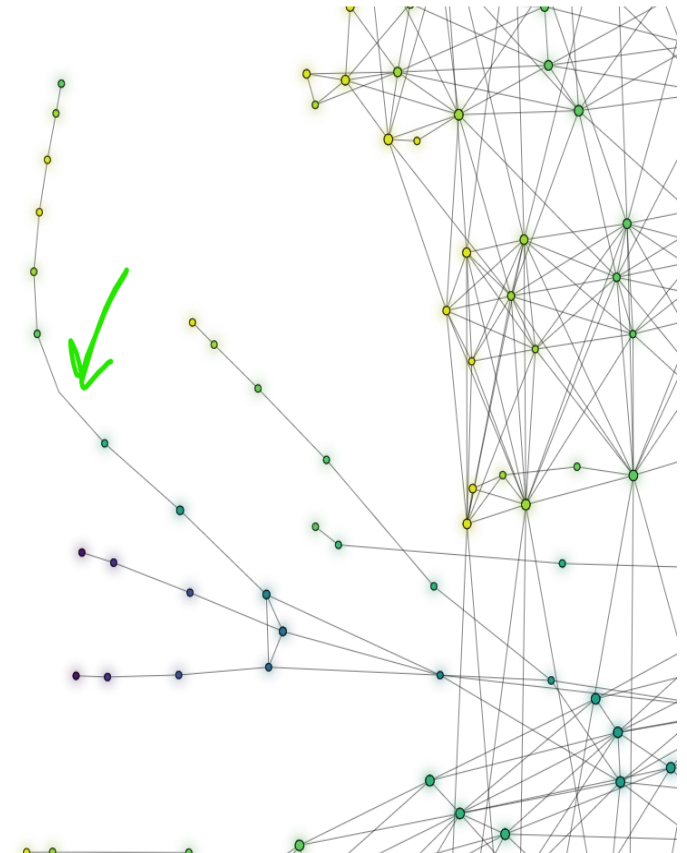
SIZE

21

MEMBERS

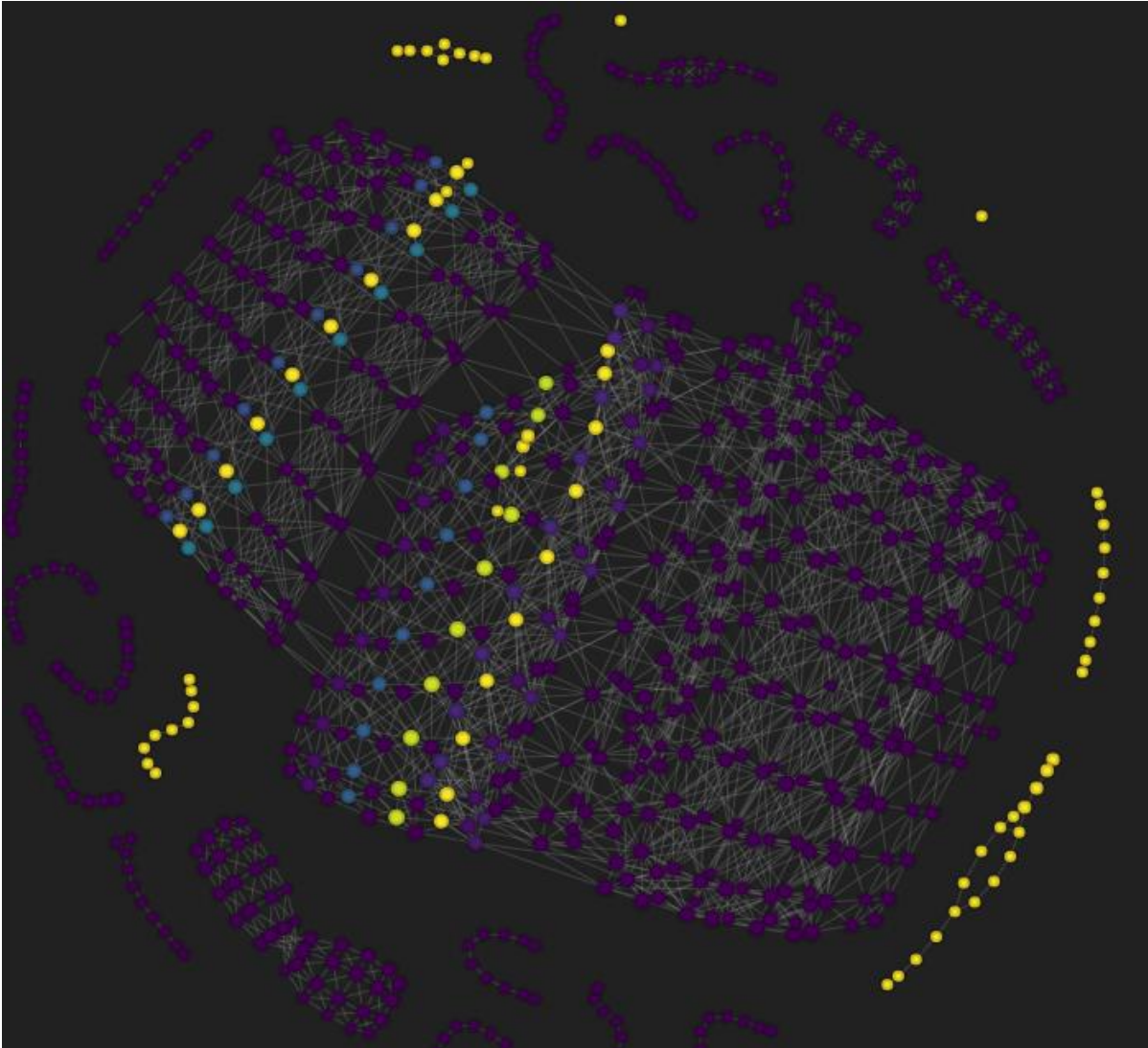
Mumbai, Maharashtra, 2021-04-03, 441475
Mumbai, Maharashtra, 2021-04-04, 452681
Mumbai, Maharashtra, 2021-04-05, 462560
Mumbai, Maharashtra, 2021-04-06, 472600
Mumbai, Maharashtra, 2021-04-07, 483042
Mumbai, Maharashtra, 2021-04-08, 491980
Mumbai, Maharashtra, 2021-04-09, 501182
Mumbai, Maharashtra, 2021-04-10, 510512
Mumbai, Maharashtra, 2021-04-11, 520498
Mumbai, Maharashtra, 2021-04-12, 527391
Mumbai, Maharashtra, 2021-04-13, 535264
Mumbai, Maharashtra, 2021-04-14, 545195
Thane, Maharashtra, 2021-04-14, 426602
Mumbai, Maharashtra, 2021-04-15, 553404

Jan to May

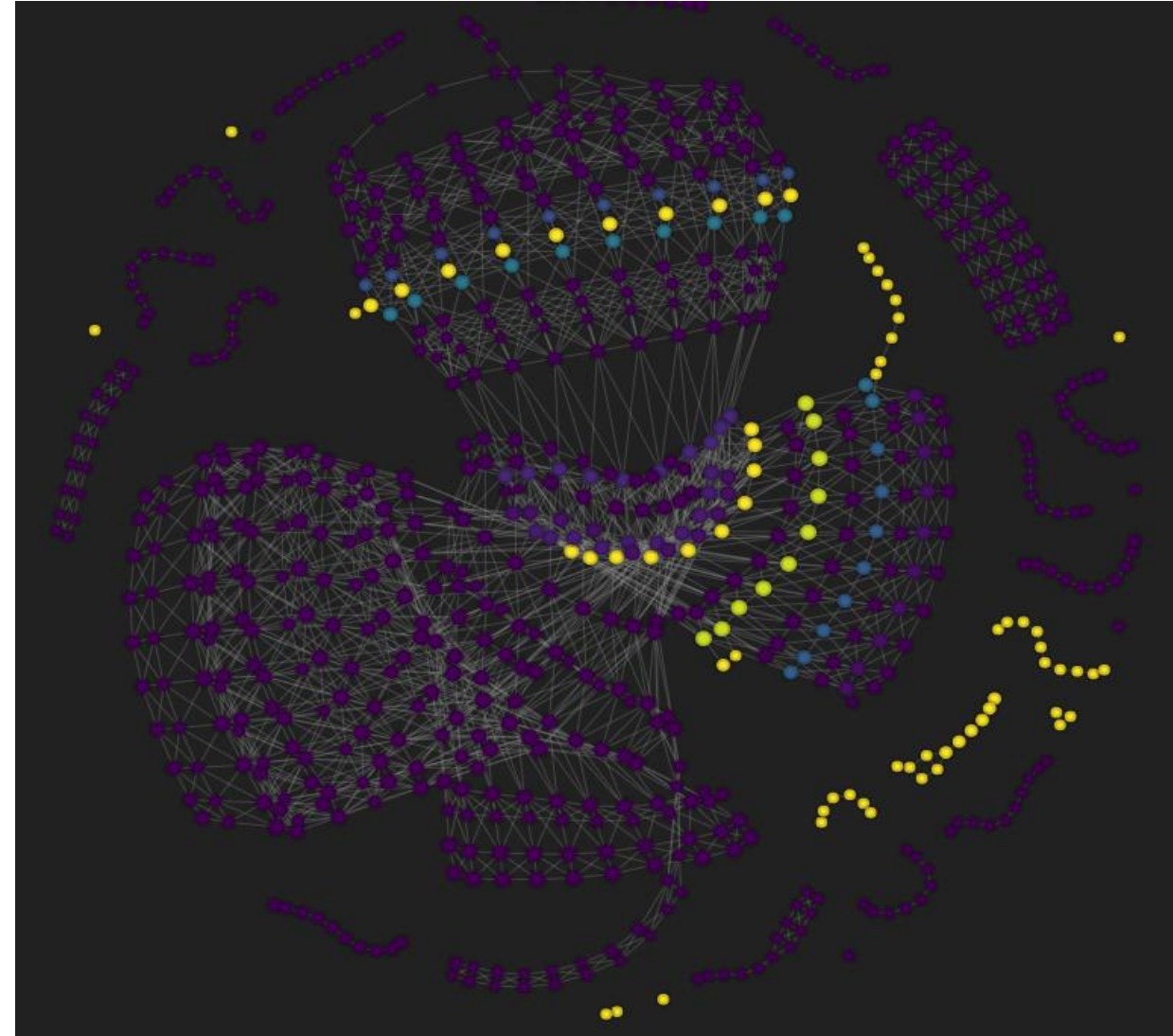


Evolution of Mapper Graph – Maharashtra Highlighted

Jan - Apr'15

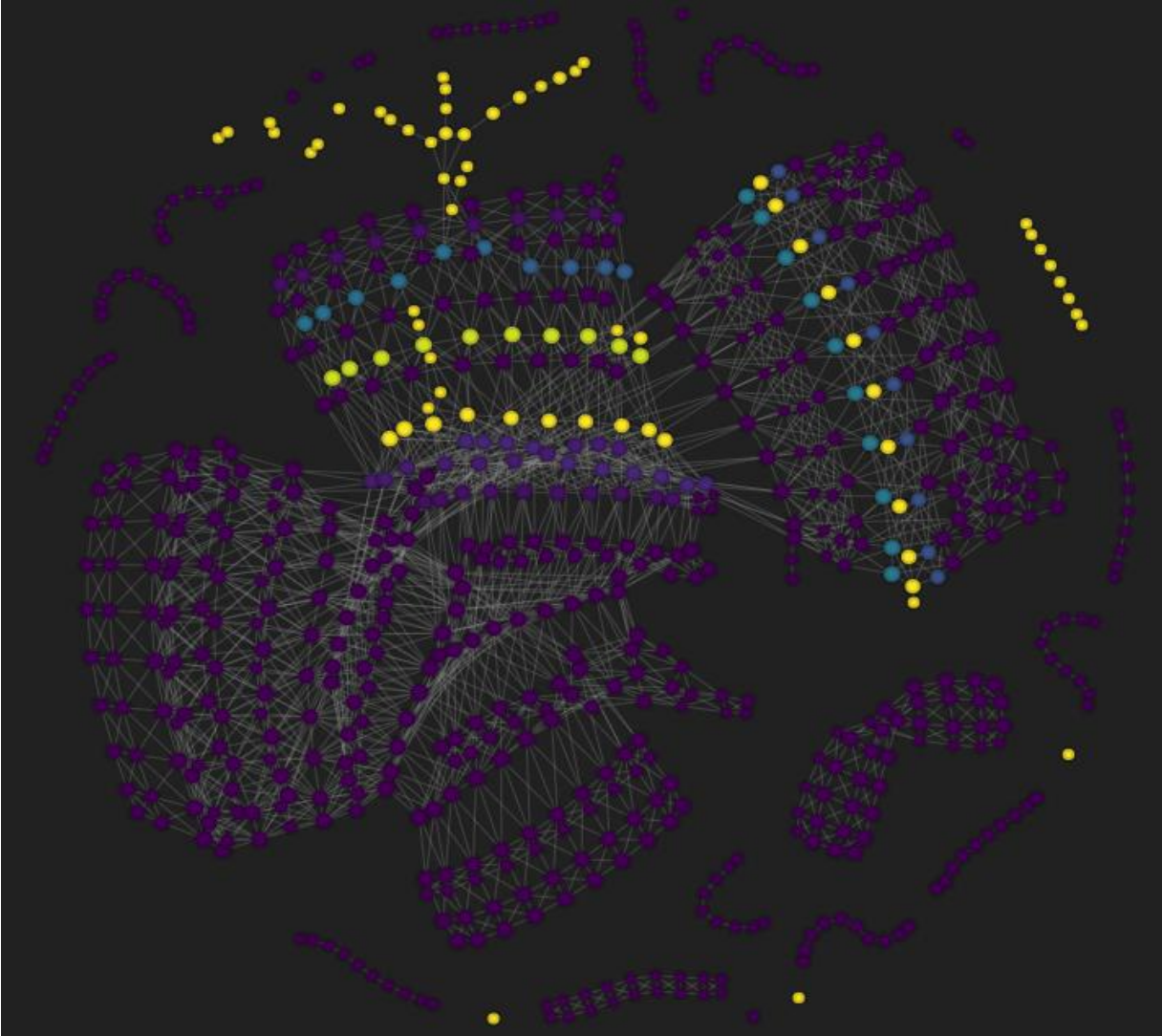


Jan – Apr'30

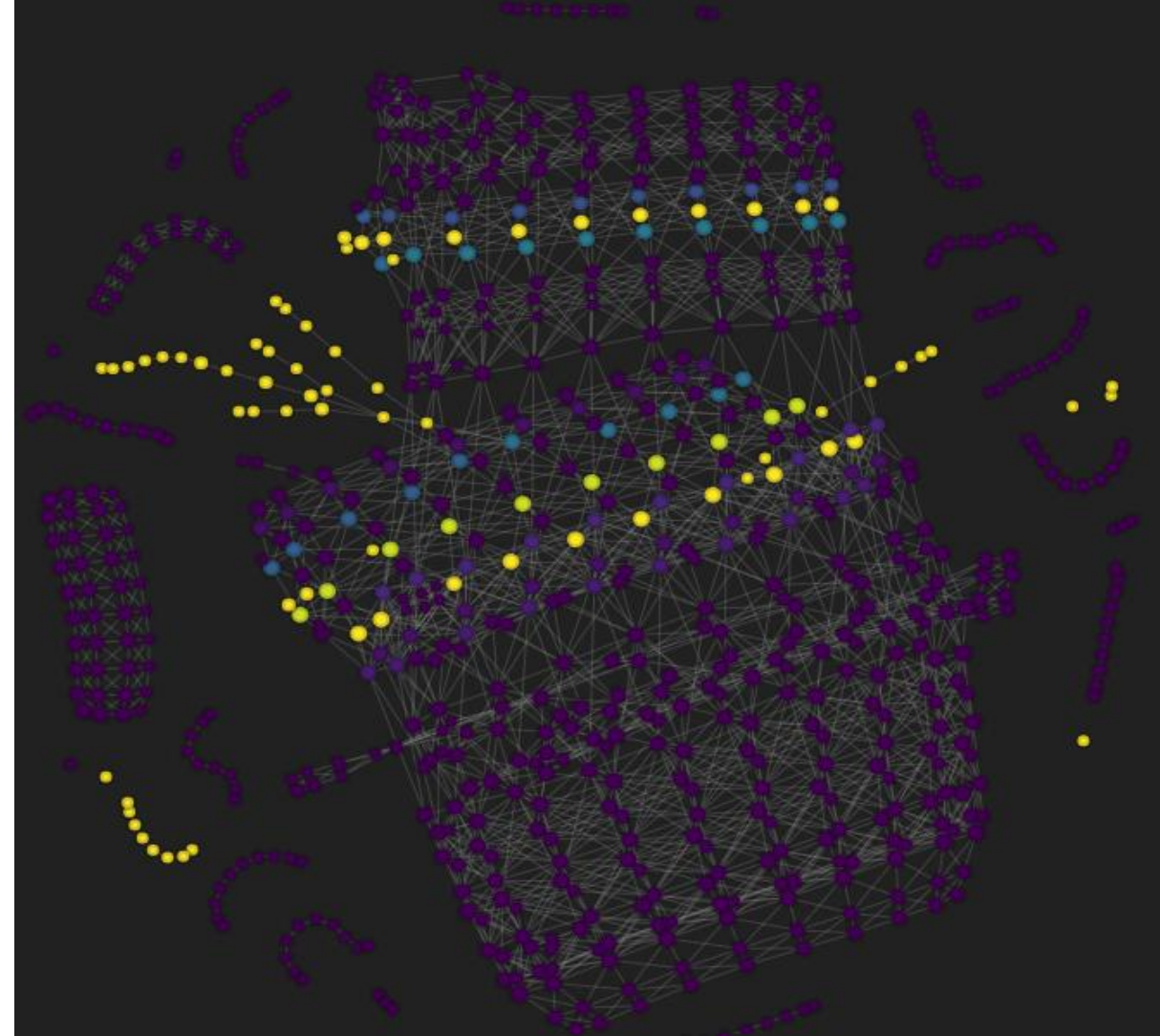


Evolution of Mapper Graph – Maharashtra Highlighted

Jan – May'16



Jan - May





Key Challenges

- Parameter Estimation: of hyperparameters for clustering algorithm
- Ensuring that insights are consistent:
 - Addressed this issue by visualizing mapper generated graphs for various time periods.
- Manually analyzing the graph to extract insights and keeping track of the same.
 - Ex: finding and keeping track of a particular state.

Critical review

- Pros:
 - Forming of flares: means outbreak in those regions
 - Corrective action can be taken.
 - Able to track gradual increase of cases over time.
- Cons:
 - Visualizations may not be consistent:
 - Depends on duration taken for visualization.
 - Does not consider the external factors like lockdown, vaccination, etc.

References

- Topological data analysis model for the spread of the coronavirus by Yiran Chen, Ismar Volić
 - doi: <https://doi.org/10.1101/2020.08.13.20174326>
- Data: COVID-19 ,INDIA
 - (<https://api.covid19india.org/>)
- A flexible Python implementation of the Mapper algorithm. Journal of Open Source Software, 4(42), 1315,
 - <https://doi.org/10.21105/joss.01315>



Queries
/suggestions

Thank You

