TBSI 清华-伯克利深圳学院 Tsinghua-Berkeley Shenzhen Institute

Exploring Dynamic Collaboration Patterns Using Zigzag Persistence

Anping Zhang

2023.10.26

Community Efforts During the COVID-19

COVID-19 Emergency Response

Business and School Re-opening Preparations

Community Bonding

Active Volunteer Force

“Shenzhen Pioneers” Volunteering Platform

Time Period

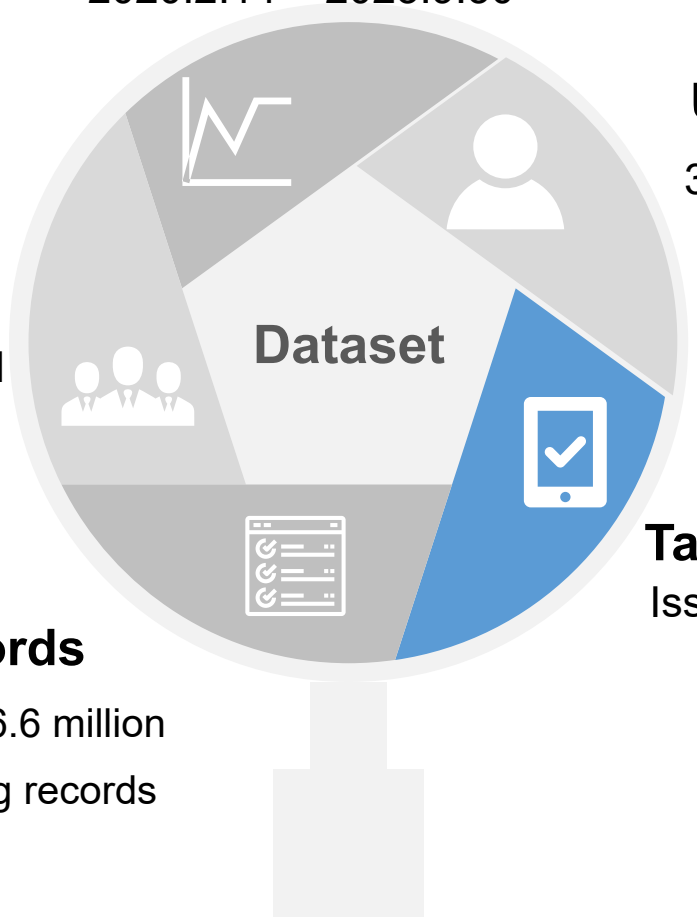
2020.2.14 – 2023.9.30

User ID

361114 Users in Total

Issuer ID

25278 Issuer in Total



Tasks

Issued 1207304 Tasks

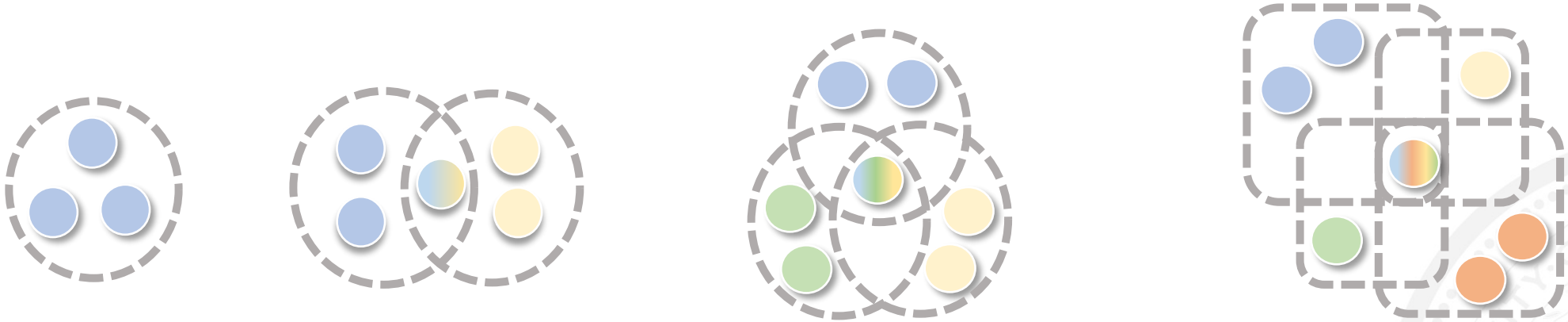
Records

More than 6.6 million volunteering records

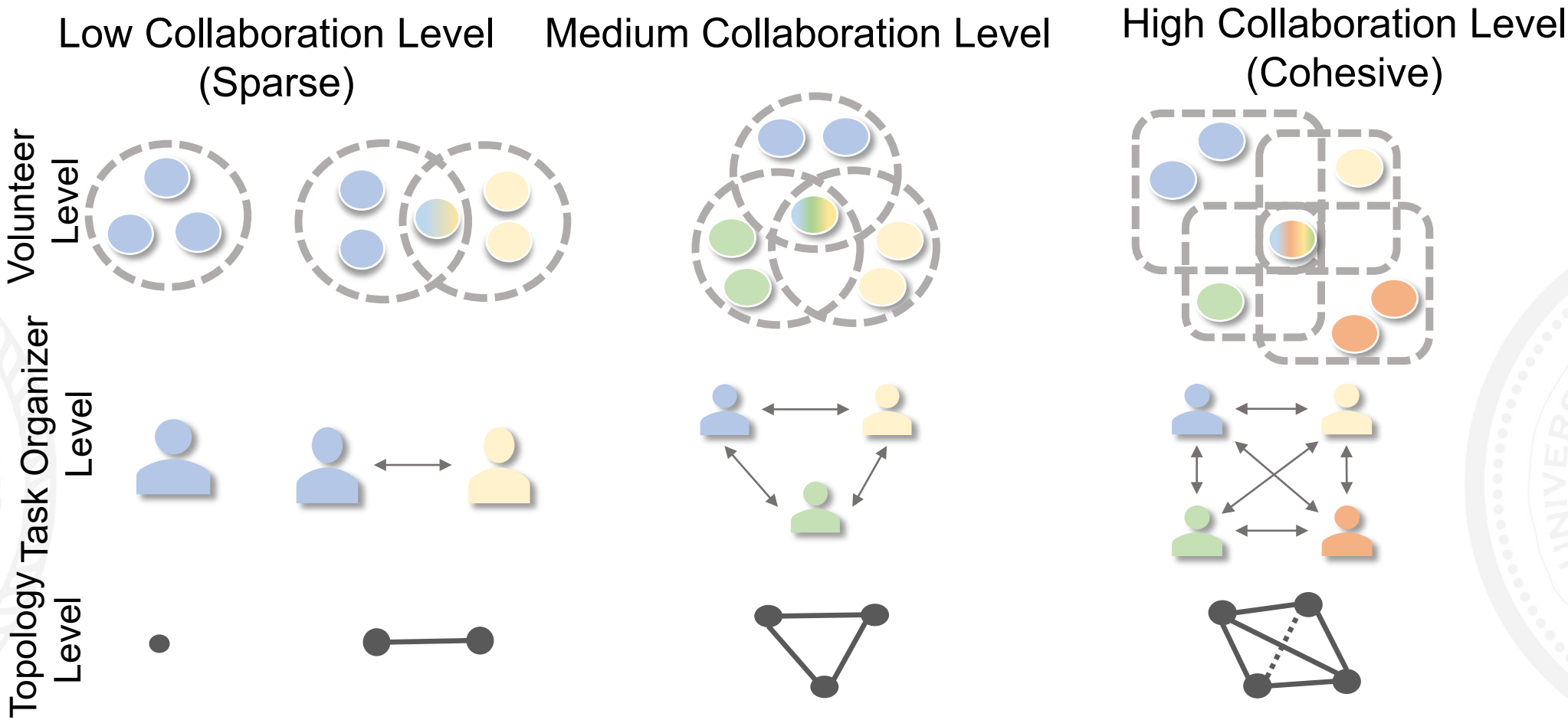
“Shenzhen Pioneers” Platform Screenshot

Zhang, Anping, et al. "Optimising self-organised volunteer efforts in response to the COVID-19 pandemic." *Humanities and Social Sciences Communications* 9.1 (2022).

Collaboration Patterns



Topology of Collaboration Patterns



Conventional graph analysis methods (e.g. centrality and connectedness) do not account for higher-dimensional structures.

Research Questions

- How to quantify dynamic collaboration patterns from a large-scale volunteering dataset?

- How to explain different collaboration patterns and changes?

Methodology

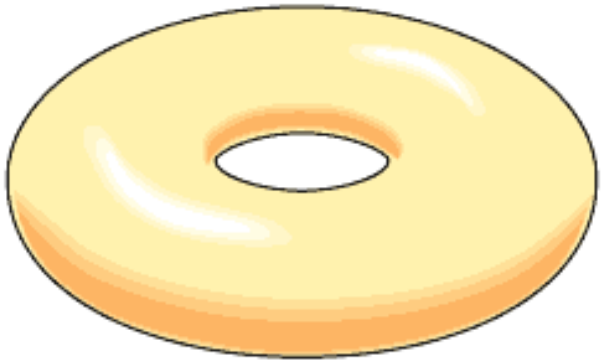
Topological data analysis

Zigzag Persistence

Zigzag Persistence-based Framework

Topological Data Analysis

- Topology is…
  - The study of holes
  - The Study of connectivity
  - Could think of it as space bending
- Betti Numbers
  - β_0 = # connected components
  - β_1 = # cycles
  - β_2 = # voids



doughnut

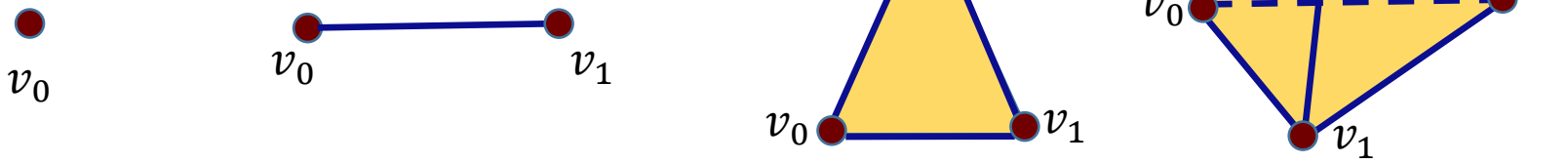
© 2006 Encyclopædia Britannica, Inc.

A topologist is a person who cannot tell the difference between a coffee mug and a donut.

—so goes a joke about a little-known scientific field crowned Tuesday with a Nobel Physics Prize

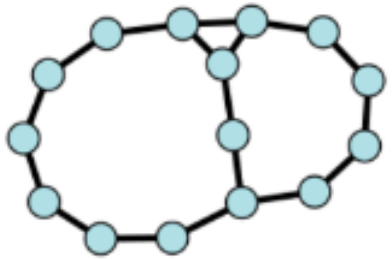
Simplicial Simplex and Complexes

- Geometric p-simplex is a convex combination of $p+1$ *(affinely) independent* points in R^N
- Complex K is a collection of simplices
  - dim(K) = highest dim of any simplex in K

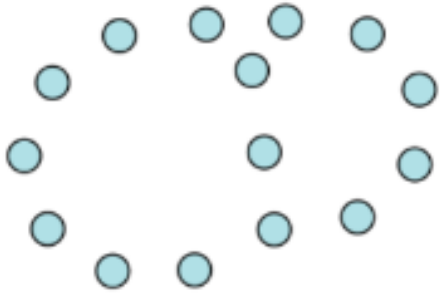


0-simplex 1-simplex 2-simplex 3-simplex

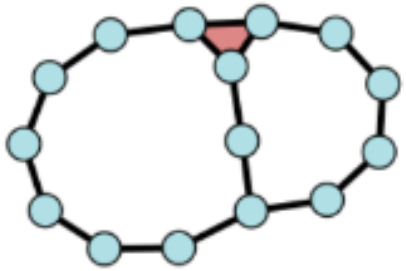
Vietoris-Rips Filtration for Graphs



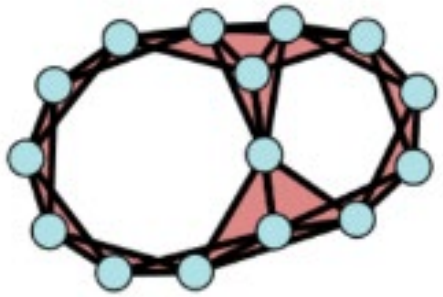
$G = (V, E)$



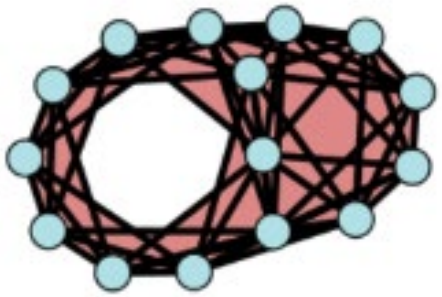
$R_0(V, d)$



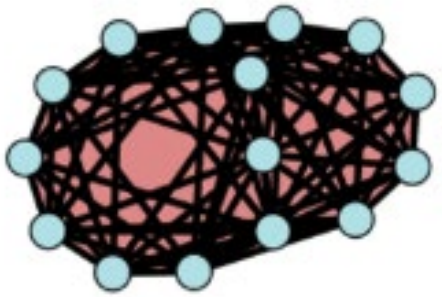
$R_1(V, d)$



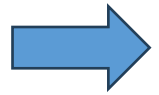
$R_2(V, d)$

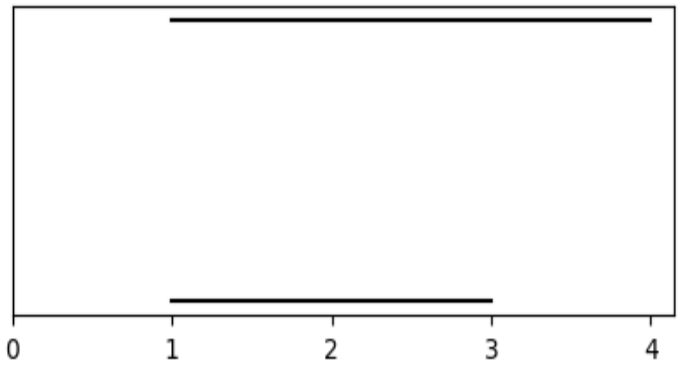


$R_3(V, d)$

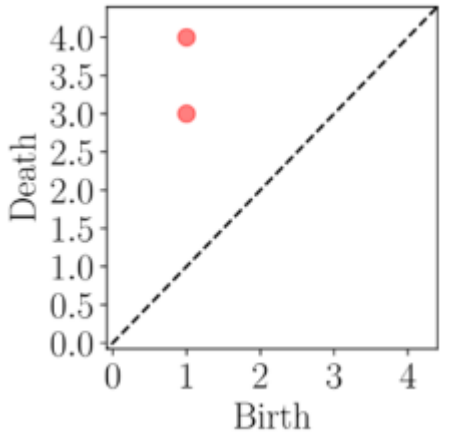


$R_4(V, d)$



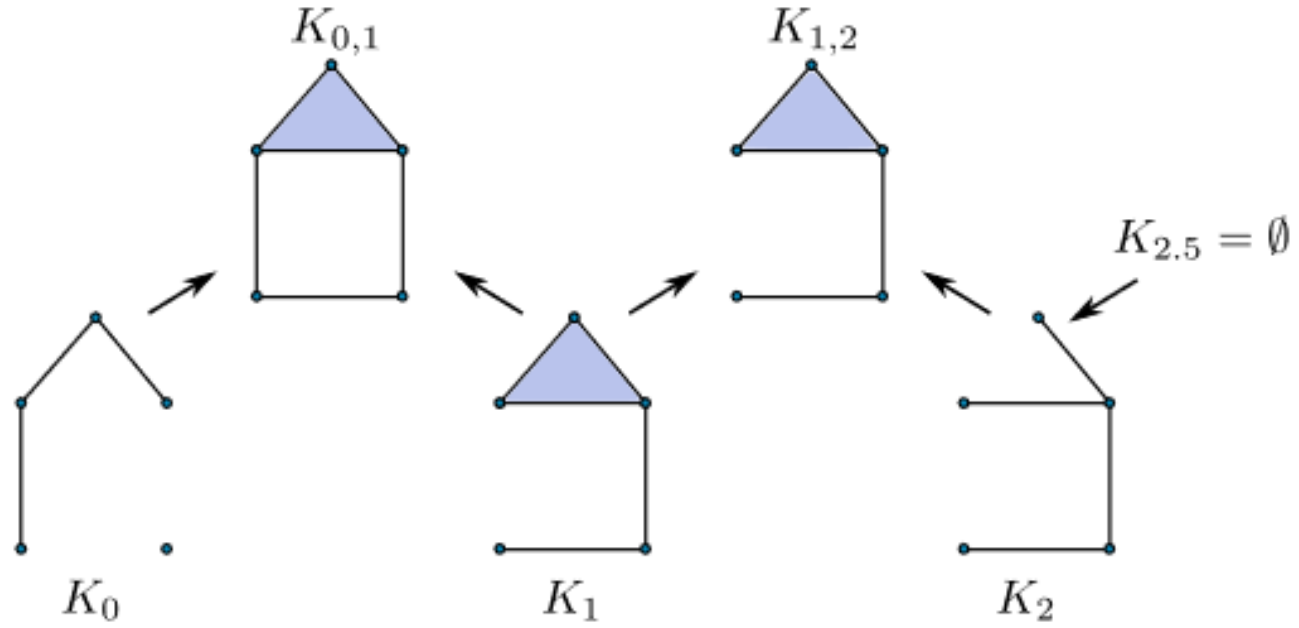


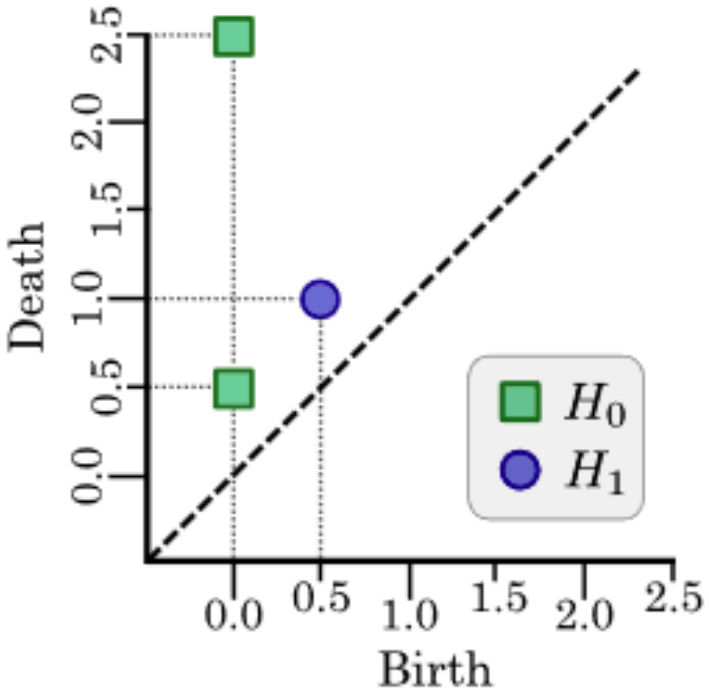
Persistence Barcode



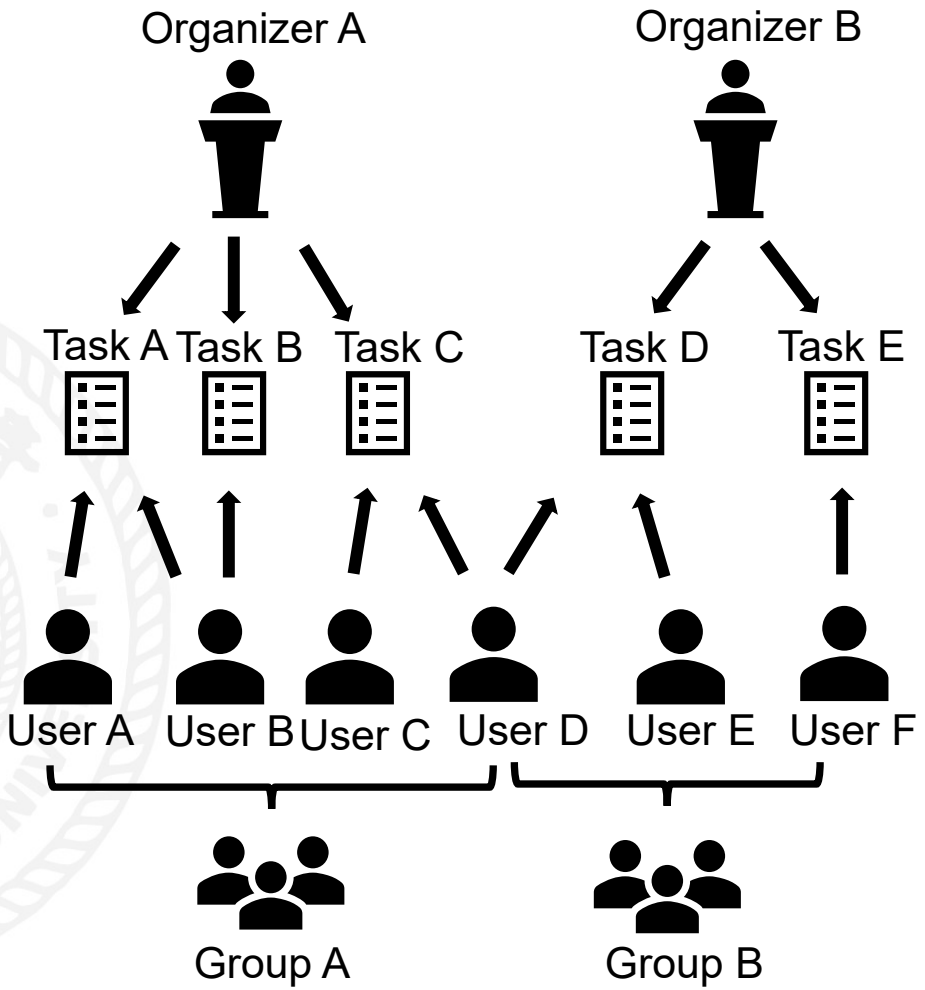
Persistence Diagram

Zigzag Persistence for Graphs

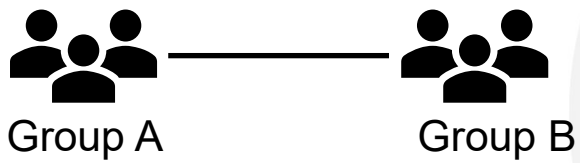




Graph Construction

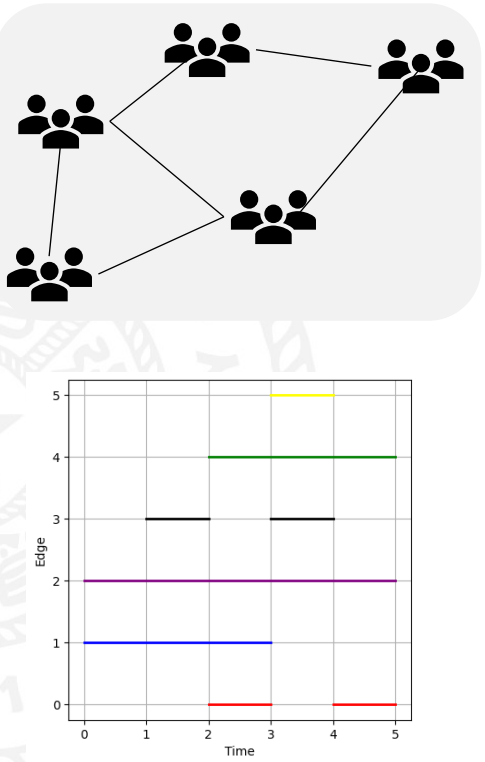


Groups are defined as a set of users who participate in tasks that are issued by the same organizer.

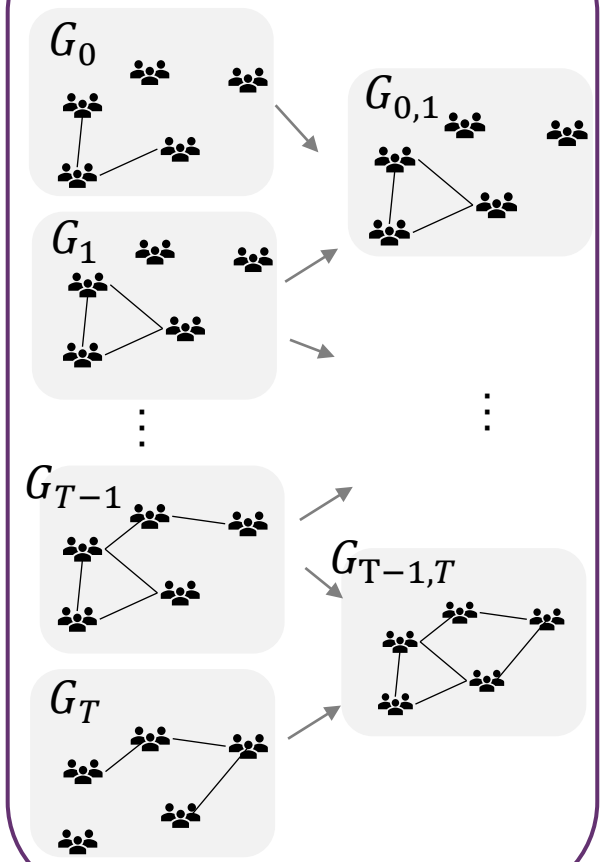


Zigzag Persistence-based Framework

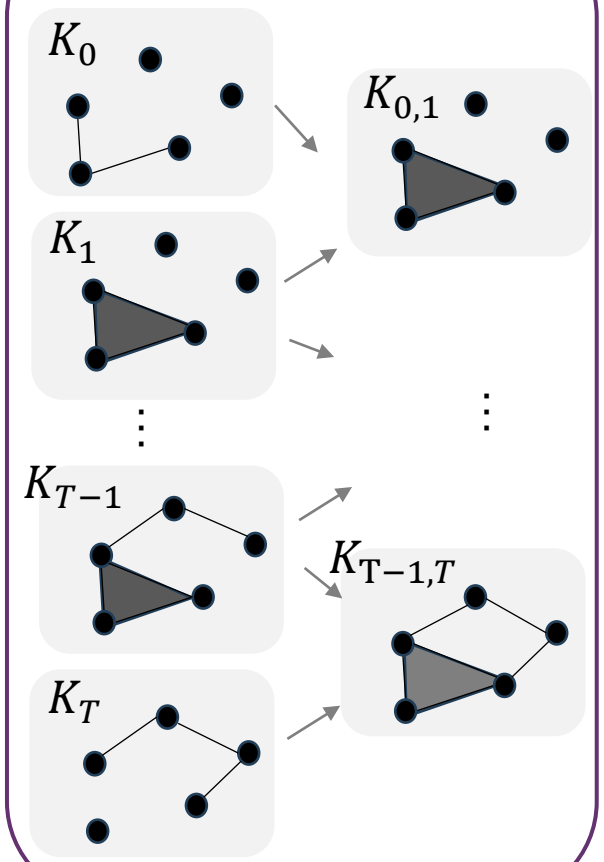
Temporal Graph



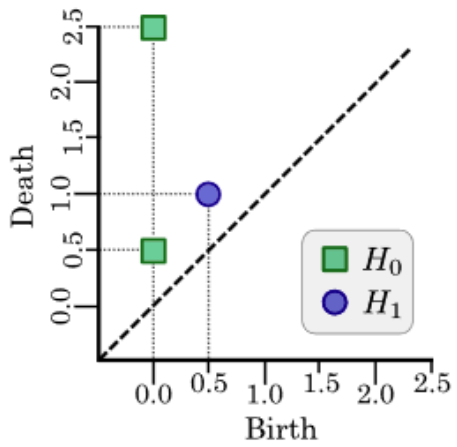
Graph Snapshots



Simplicial Complexes



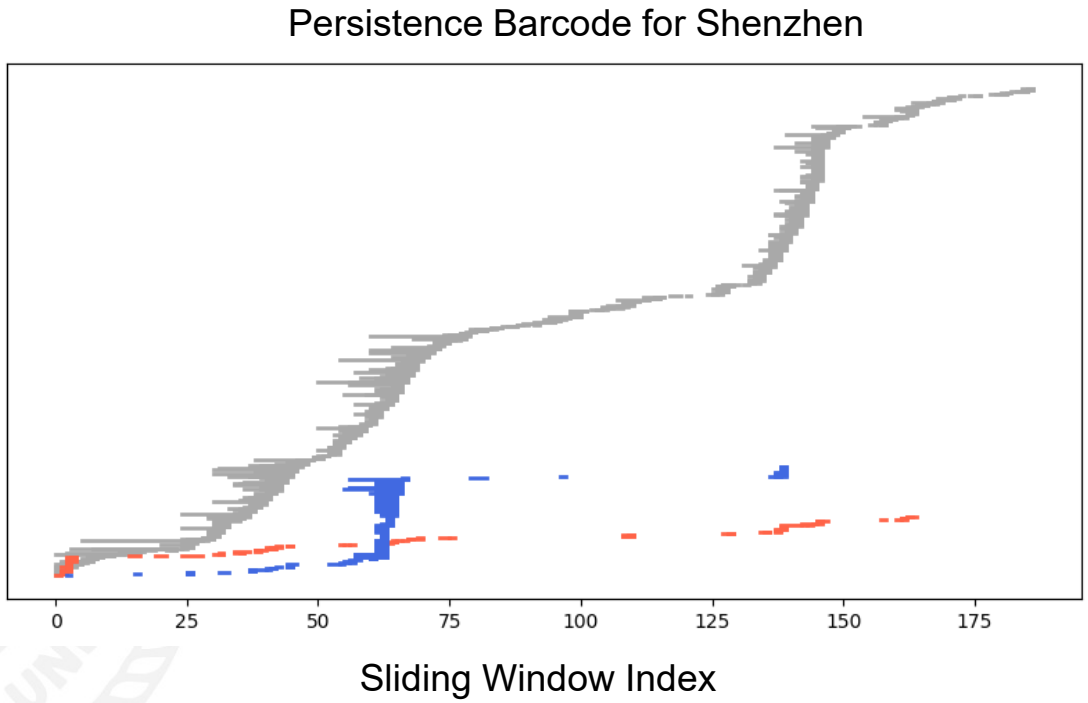
Zigzag Persistence

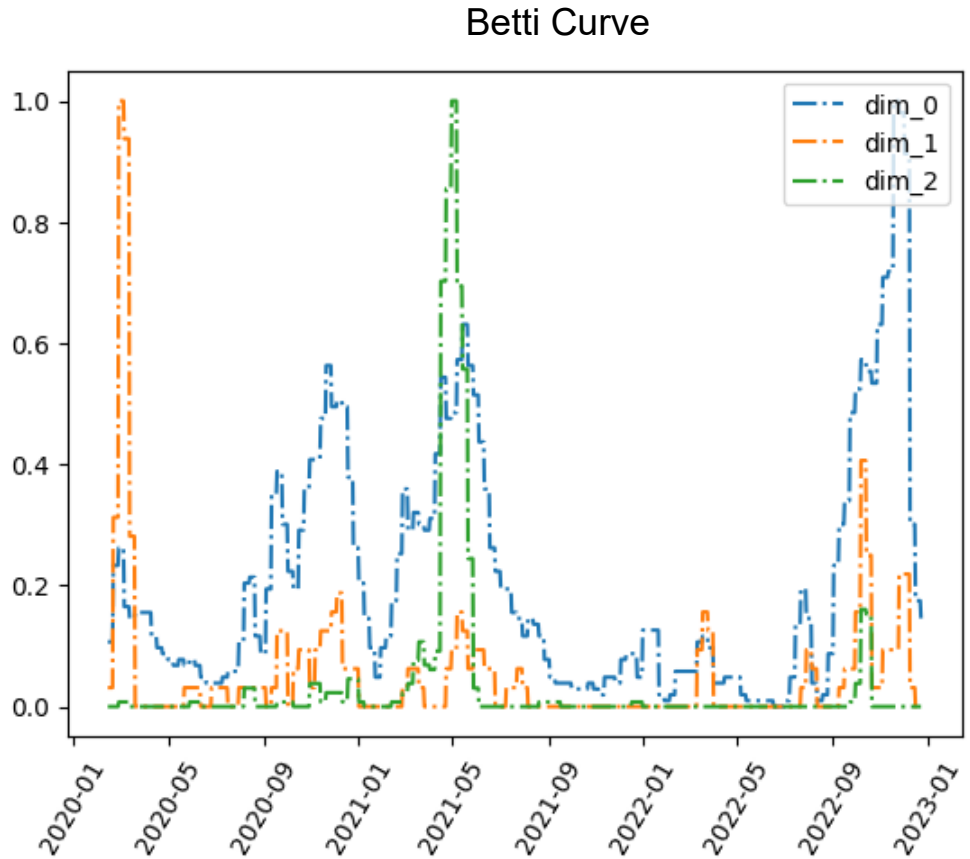


Experimental Results

City, District, and Street Level

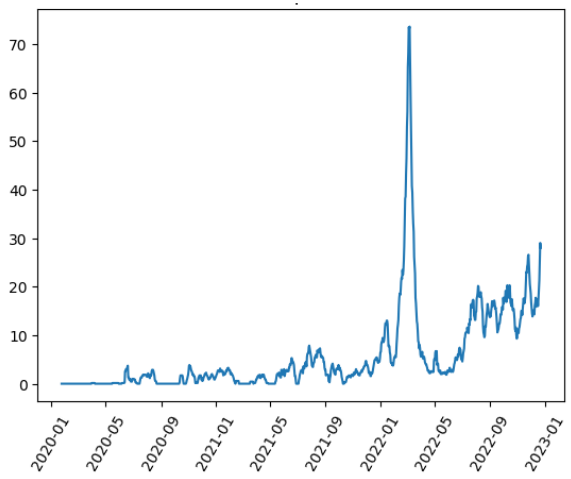
City-level: Betti Curve



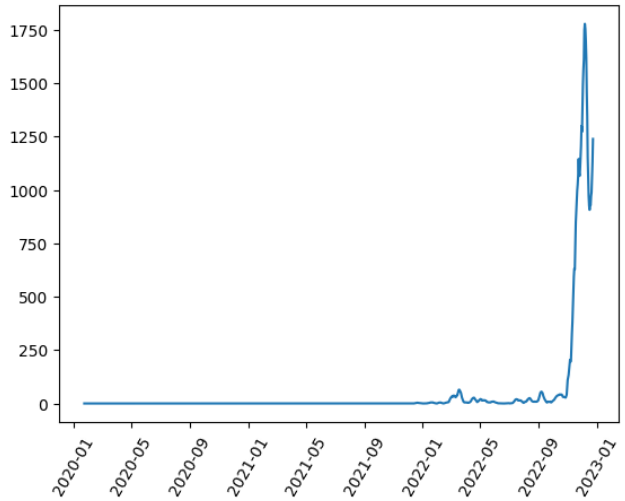


City Level: Pandemic Influence

Guangdong_covid_import

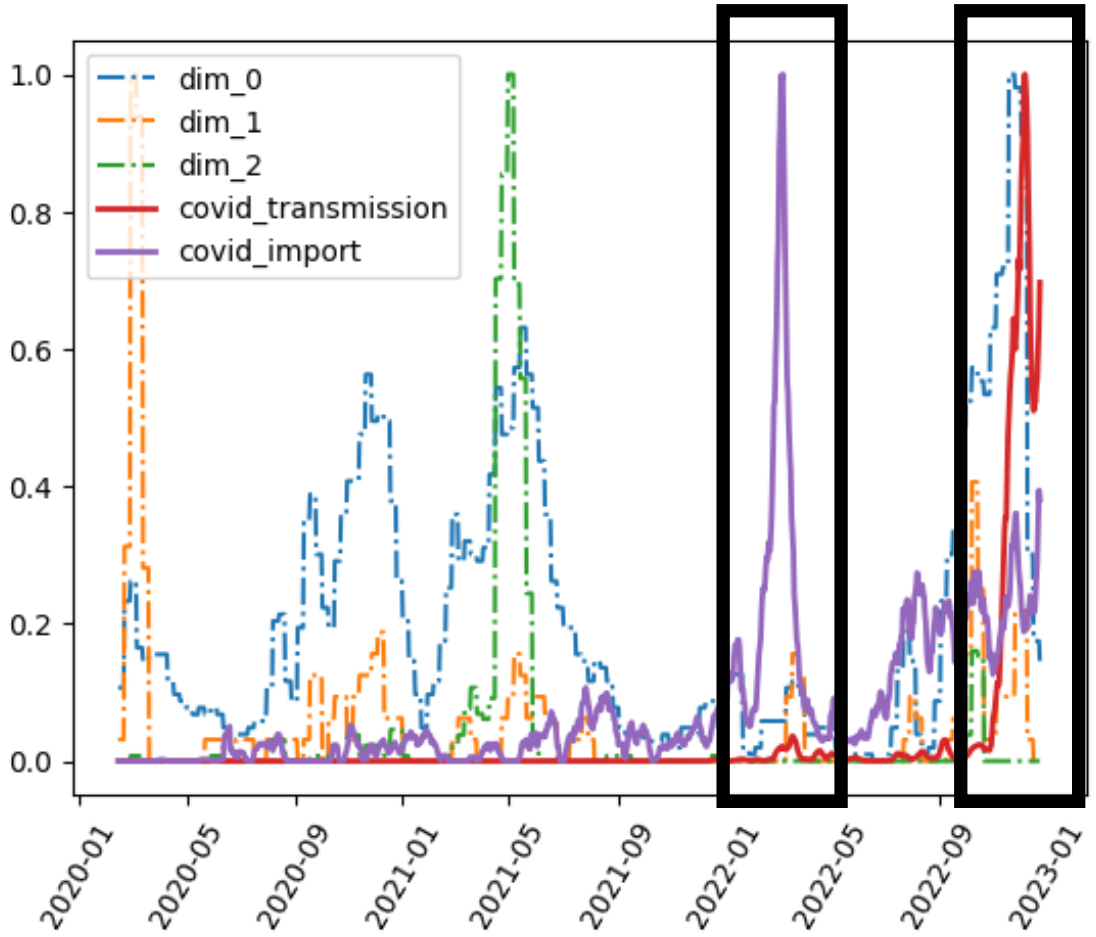


Guangdong_covid_trans



Pearson Correlation T-test (0.05)

	covid_trans	covid_import
dim_0	0.479 (6.55e-61)	0.106(0.001)
dim_1	0.110 (3.93e-4)	0.0007(0.983)
dim_2	-0.054 (0.0783)	-0.107(5.14e-4)



District Level: Normalized Conditional Entropy (NCE)

- **P-NCE**: Uncertainty on Users (U)' Participation Rate($P(u)$)

$$\hat{H}(X \mid U) = \frac{H(X \mid U)}{\sum_{u \in \mathcal{U}} H(X \mid U = u)} = \frac{-\sum_{u \in \mathcal{U}} \Pr(U = u)(P(u)\log(P(u)) + (1 - P(u))\log(1 - P(u))}{-\sum_{u \in \mathcal{U}} (P(u)\log(P(u)) + (1 - P(u))\log(1 - P(u))}$$

- **O-NCE**: Uncertainty on Choosing a Task Organizer(O)

$$\hat{H}(O \mid U) = \frac{H(O \mid U)}{\sum_{u \in \mathcal{U}} H(O \mid U = u)} = \frac{-\sum_{u \in \mathcal{U}} \Pr(U = u) \sum_{o \in \mathcal{O}} O(o,u)\log(O(o,u))}{-\sum_{u \in \mathcal{U}} \sum_{o \in \mathcal{O}} O(o,u)\log(O(o,u))}$$

District Level: Organizational Speed (η)

Double Exponential Model

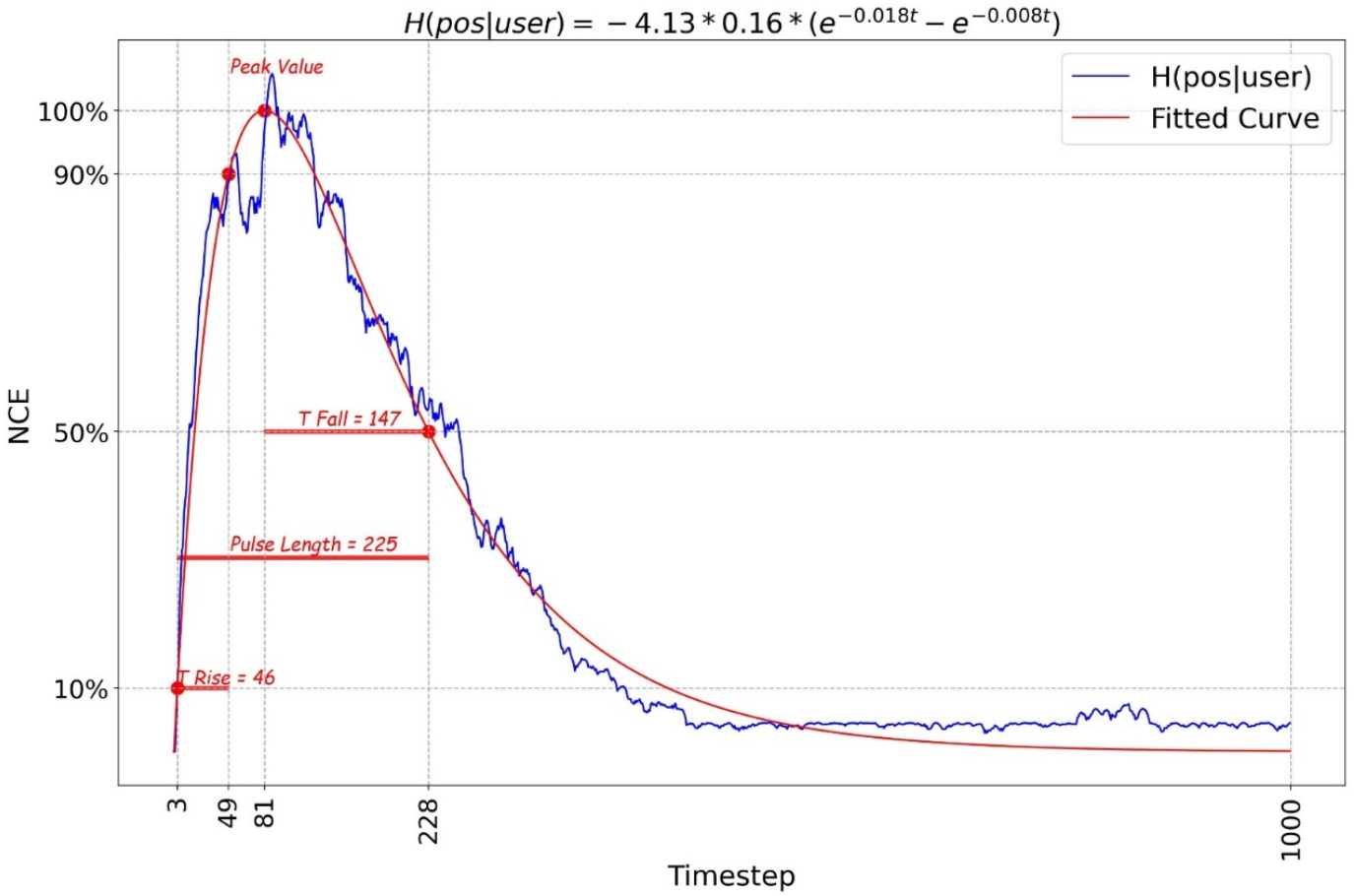
$$NCE(t) = A * I * (e^{-\alpha t} - e^{-\beta t})$$

Where I is the peak value, $A = f(\alpha, \beta)$:

$$A(\alpha, \beta) = \frac{1}{e^{-\alpha \frac{\ln(\beta)-\ln(\alpha)}{(\beta-\alpha)}} - e^{-\beta \frac{\ln(\beta)-\ln(\alpha)}{(\beta-\alpha)}}}$$

Organizational Speed

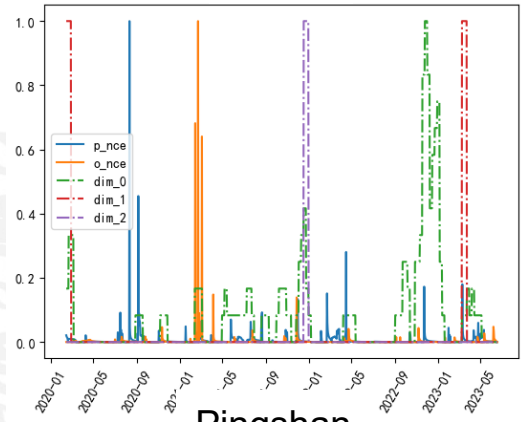
$$\eta = \frac{T_{\text{Half}}}{T_{\text{Fall}}} \approx \frac{\frac{1}{\alpha} - \frac{\ln(\beta)-\ln(\alpha)}{\beta-\alpha}}{n - \frac{\ln(\beta)-\ln(\alpha)}{\beta-\alpha}}$$



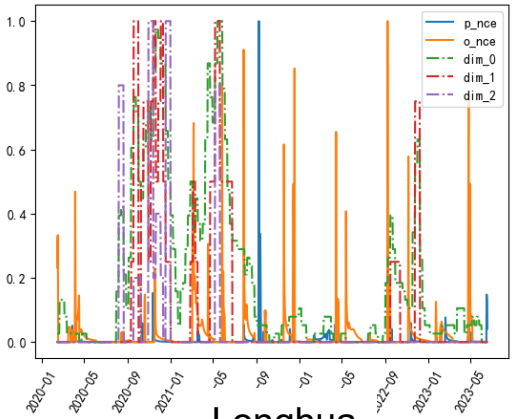
District Level: Self-Organization Influence

- Motivation: why did districts show different collaboration levels?

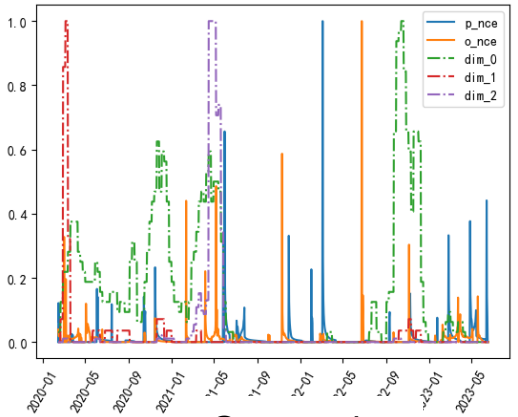
Luohu



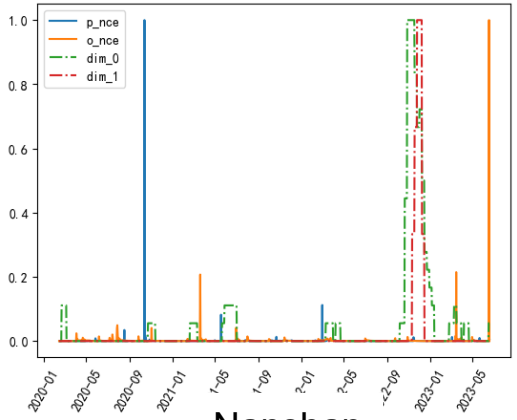
Longgang



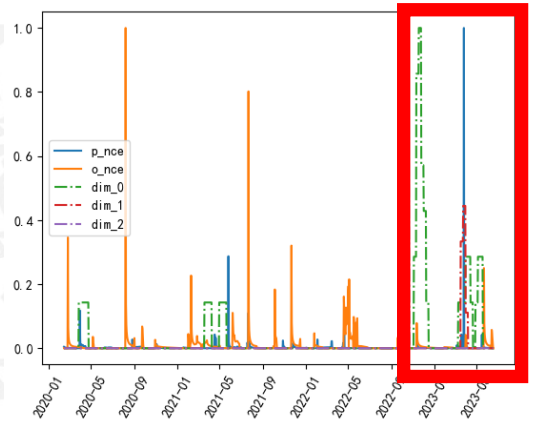
Futian



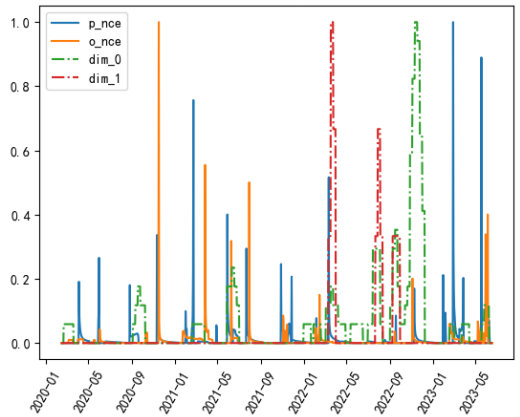
Baoan



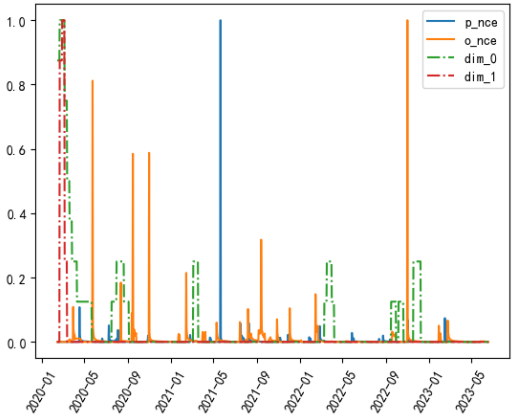
Pingshan



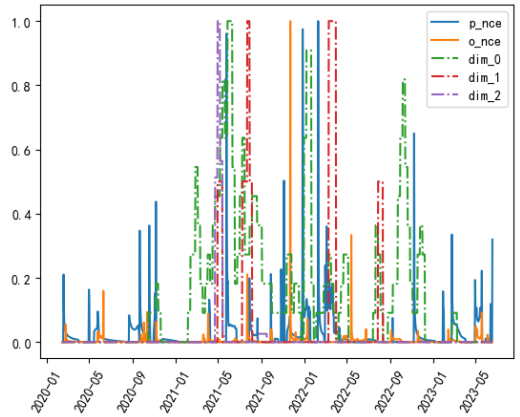
Longhua

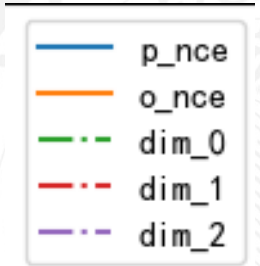


Guangming



Nanshan





District Level: Self-Organization Influence

Pearson Correlation T-Test (Correlation(P-Value))

	P-NCE&dim0	P-NCE&dim1	P-NCE&dim2	O-NCE&dim0	O-NCE&dim1	O-NCE&dim2
Guangming	-0.017(0.548)	-0.01(0.722)	Na	-0.031(0.276)	-0.018(0.520)	Na
Longhua	0.025(0.374)	0.016(0.575)	Na	0.019(0.518)	-0.032(0.261)	Na
Longgang	-0.028(0.323)	-0.013(0.662)	-0.007(0.811)	0.029(0.304)	0.001(0.972)	-0.058(0.042)
Nanshan	0.075(0.008)	0.083(0.004)	-0.046(0.105)	0.013(0.659)	0.017(0.543)	-0.01(0.727)
Luohu	0.017(0.564)	0.045(0.113)	-0.014(0.627)	0.028(0.335)	-0.013(0.647)	-0.007(0.795)
Pingshan	0.032(0.267)	0.11(0.0001)	-0.01(0.72)	-0.02(0.484)	-0.033(0.250)	-0.025(0.378)
Futian	-0.054(0.058)	-0.031(0.284)	-0.031(0.282)	0.015(0.60)	0.067(0.018)	0.086(0.002)
Baoan	-0.009(0.756)	-0.004(0.885)	Na	0(0.99)	-0.011(0.70)	Na

Street Level: Point-of-interest Influence

Point-of-interest of 72 Streets

POI Type	Yuanling Street	Pinghu Street	…
Science/Culture & Education Service	256	662	…
Transportation Service	267	900	…
Governmental Organization & Social Group	103	462	…
Tourist Attraction	8	61	…
Medical Service	90	876	…
Residential Area	25	187	…
Accommodation Service	19	550	…
Daily Life & Sports & Recreation Service	882	2292	…
Industrial Park & Business Office Building	38	339	...

Street Level: Point-of-interest Influence

Definition 3.1. The **persistence statistics** vector of $\mu : B \to \mathbb{Z}_{>0}$ consists of:

1) the mean, the standard deviation, the median, the interquartile range, the full range, the 10^{th}, 25^{th}, 75^{th} and 90^{th} percentiles of the births p, the deaths q, the midpoints $\frac{p+q}{2}$ and the lifespans $q-p$ for all intervals $[p,q]$ in B counted with multiplicity;
2) the total number of bars (again counted with multiplicity), and
3) the *entropy* of μ, defined as the real number

$$E_\mu := -\sum_{[p,q]\in B} \mu_{p,q} \cdot \left(\frac{q-p}{L_\mu}\right) \cdot \log\left(\frac{q-p}{L_\mu}\right),$$

where L_μ is the weighted sum

$$L_\mu := \sum_{[p,q]\in B} \mu_{p,q} \cdot (q-p). \quad (1)$$

Persistence Feature Y ($72 * 38$) $=$ POI Matrix X ($72 * 38$) $*$ β ($9 * 38$)

Regression Coefficients for Different POI Types

POI Type	Coefficient(β_0)	Coefficient (β_1)	Coefficient (β_2)
Science/Culture & Education Service	0.377	0.243	0.570
Transportation Service	0.176	0.135	0.079
Governmental Organization & Social Group	0.264	0.479	0.167
Tourist Attraction	0.138	0.084	0.107
Medical Service	0.228	0.165	0.184
Residential Area	0.520	0.278	0.277
Accommodation Service	0.09	0.188	0.079
Daily Life & Sports & Recreation Service	0.264	0.232	0.084
Industrial Park & Business Office Building	0.347	0.529	0.262

Conclusions

Conclusions & Future Work

Conclusion & Future Works

- Conclusion
  - Quantify:
    - We proposed a zigzag persistence-based framework that can quantify collaboration patterns.
  - Explain:
    - We explained collaboration patterns at the city, district, and street levels, the results indicate that the **pandemic**, **organizational levels**, and **regional points of interest** individually influence volunteer collaborations, respectively.
- Future Works
  - Implement the framework on other dynamic datasets to verify model generalization.

Fin