



TBSI

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

# Career Analysis by Spatial-Temporal Network

Shi Mao

Yuan Wang

Yanfeng Chen

# Outline



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

- Introduction
- Data Crawling and Parsing
- Spatial-Temporal Graph Construction & Visualization
- Data Analysis
- Result and Conclusion
- Team Contribution



TBSI

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

# Introduction

- Motivation

Trying our best by applying learning algorithm to help people to find their suitable job.

- Problem

Conventional algorithms neglect the relationship of users' mentors, schoolmates and colleagues.



TBSI

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

# Data Crawling and Parsing

- Data Crawling

- From a public government website

- <http://www.sz.gov.cn/cn/xxgk/zfxxgj/sldzc/sz97404/crg/>

- All Shenzhen government officer resumes



TBSI

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

# Data Crawling and Parsing

## 深圳市委副书记，深圳市人民政府市长、党组书记

陈如桂，男，1962年9月生，汉族，广东廉江人，1992年10月加入中国共产党，1983年8月参加工作，研究生学历，工学博士，高级工程师，享受国务院特殊津贴专家。

1979.09—1983.08 桂林冶金地质学院物探系地球物理勘探专业大学本科学习

1983.08—1986.08 桂林冶金地质学院教师

1986.08—1989.05 中南工业大学地质系应用地球物理专业硕士研究生学习

1989.05—1991.01 广州市建筑科学研究所干部

1991.01—1992.09 广州市建筑科学研究所测试技术研究室副主任

1992.09—1992.10 广州市建筑科学研究所测试技术研究室主任

1992.10—1998.03 广州市建筑科学研究所副所长（其间：1995.09—1998.02中南工业大学地质系应用地球物理专业博士研究生学习）

1998.03—1999.08 广州市建筑科学研究院院长、市建筑集团有限公司副总工程师

1999.08—1999.12 广州市建筑集团有限公司副董事长、副总经理

1999.12—2000.06 广州市建筑集团有限公司党委副书记、副董事长、副总经理

2000.06—2001.07 广州市建筑集团有限公司董事长、总经理、党委副书记

2001.07—2003.05 广州市建设委员会副主任（广州正局级）

2003.05—2007.03 广州市建设委员会主任、工委书记

2007.03—2010.06 广州市人民政府秘书长、党组成员、办公厅党组书记

2010.06—2011.12 广州市委常委、秘书长

2011.12—2015.09 广州市委常委，广州市人民政府副市长

2015.09—2016.09 广州市委副书记、政法委书记

2016.09—2017.01 中山市委书记、市人大常委会主任候选人

2017.01—2017.07 中山市委书记、市人大常委会主任

2017.07—2017.08 深圳市委副书记，市人民政府副市长、代理市长、党组书记

2017.08—今 深圳市委副书记，市人民政府市长、党组书记

广东省第十二届省委委员



# Data Crawling and Parsing

- Data Parsing
  - Using a public Github repository  
<https://github.com/hankcs/pyhanlp>
  - Manual adjustment



TBSI

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

## 李春生

时间：1979—1983 地点：河南省 机构：河南大学历史系历史专业 职位：学习

时间：1983—1985 地点：河南省 机构：郑州市第七中学 职位：教师

时间：1985—1991 地点：河南省 机构：共青团 职位：省委干事 副科长 秘书

时间：1991—1996 地点：河南省 机构：省委学校部，维权办，希望工程办 职位：副部长，副主任，副主任

时间：1996—1998 地点：河南省 机构：共青团 职位：省委少年部

时间：1998—1999 地点：河南省新县 机构：县委 职位：副书记

时间：1999—2000 地点：河南省新县 机构：县委，县政府 职位：副书记，副县长

时间：2000—2002 地点：河南省新县 机构：县委，县政府 职位：副书记 县长，党组书记

时间：2002—2003 地点：河南省新县 机构：县委 职位：书记

时间：2003—2004 地点：河南省信阳市 机构：市委，政法委 职位：常委，书记

时间：2004—2006 地点：河南省公安厅 机构：党委，政治部 职位：委员，主任

时间：2006—2008 地点：河南省公安部政治部 机构：人事训练局 职位：局长

时间：2008—2013 地点：河南省 机构：公安部政治部，人事训练局 职位：副主任，局长

时间：2013—2015 地点：广东省 机构：广东省政府，省公安厅，省委政法委 职位：副省长 党组成员，厅长 党委书记 督察长，副书记

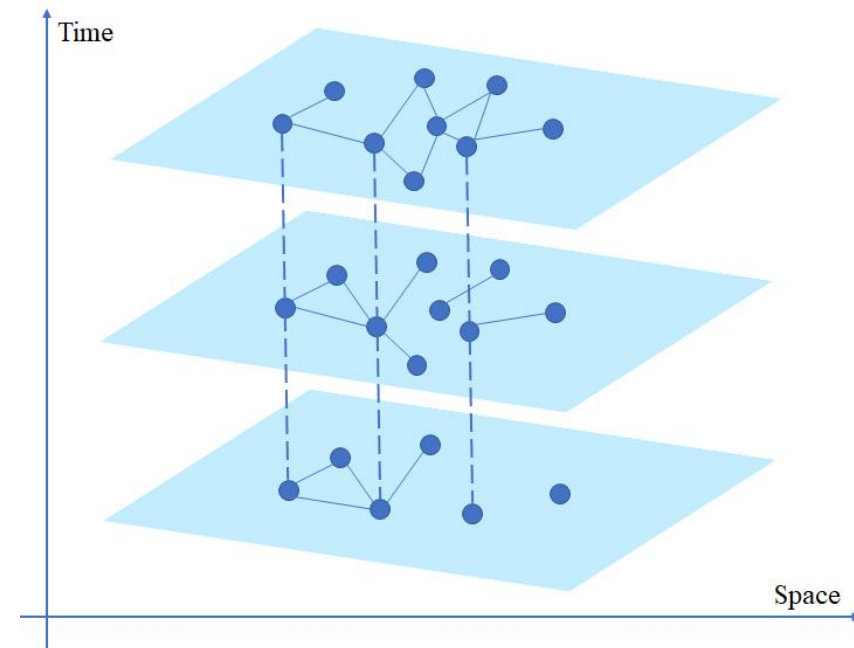
时间：2015—今 地点：广东省 机构：省政府，省公安厅，省委政法委 职位：副省长 党组成员，厅长 党委书记 督察长，第一副书记





# Spatial-Temporal Graph

- Spatial-temporal networks are spatial networks whose topology and parameters change with time.
- Specifically for Social Network, we consider the change of social connection through time
- Construction considerations:
  - How to establish connection?
  - How to choose time step?
  - Labeling

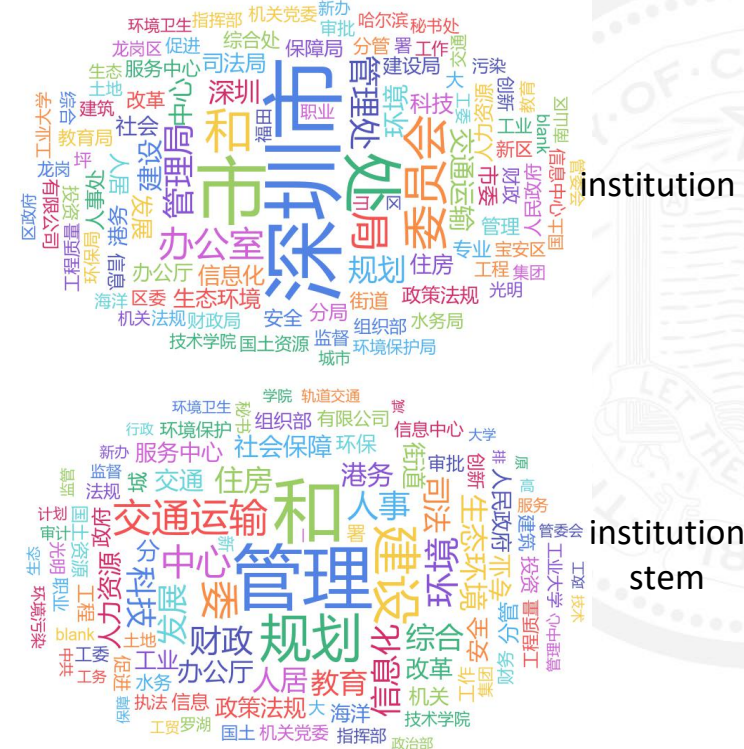




# Spatial-Temporal Graph: Connection

- Connected if serving the same institution?
  - too Sparse: 108 repeated in all 720 institutions
- Measure the distance of nodes by the **stem** of their institution 'similarity'.
  - By fuzzy matching

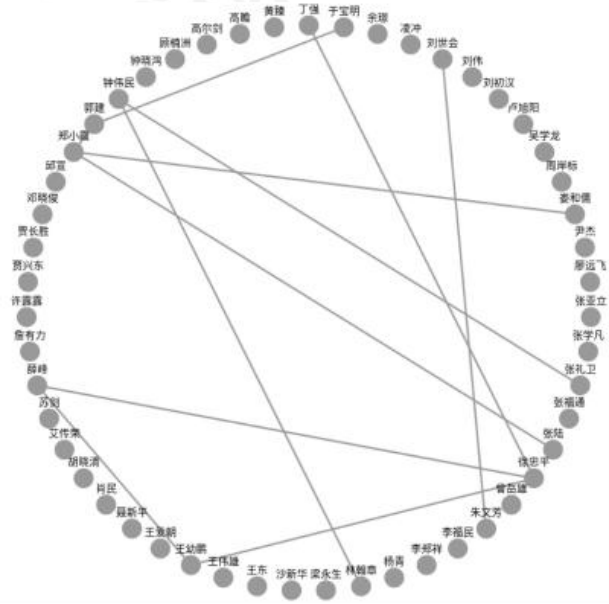
| name                | institution                              | institution stem             |
|---------------------|--|------------------------------|
| 于宝明<br>Baoming Yu   | 深圳市交通局<br>Shenzhen transportation Bureau | 交通<br>trasnportation         |
| 徐忠平<br>Zhongping Xu | 智慧交通处<br>Intel-transportation Division   | 智慧交通<br>intel-transportation |
| 张福通<br>Futong Zhang | 深圳市财政局<br>Shenzhen Finance Bureau        | 财政<br>Finance                |



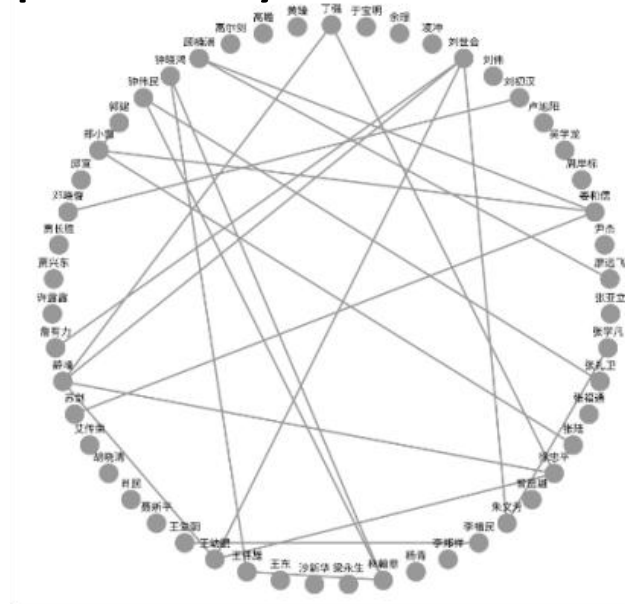


# Spatial-Temporal Graph: Connection

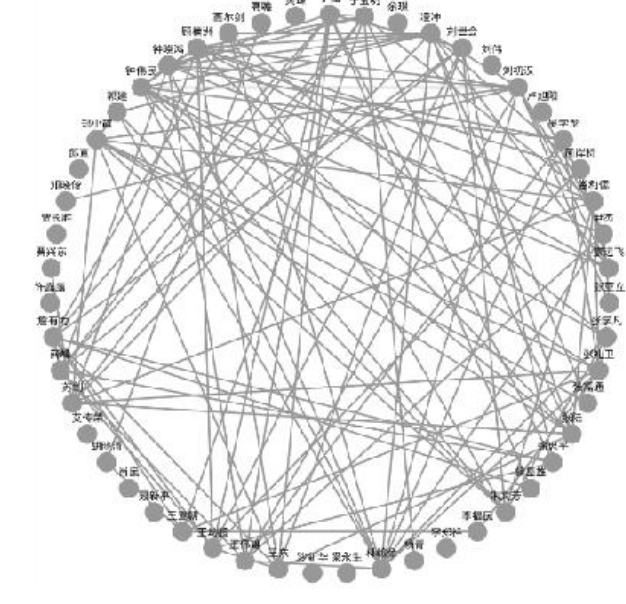
- Build the graph by epsilon-neighborhood
  - different people might have different number of neighbors
- Fine-tune the epsilon by visualization



threshold 0



threshold 20

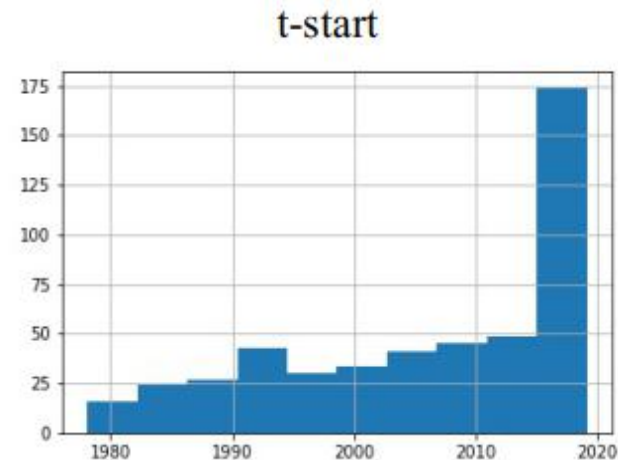
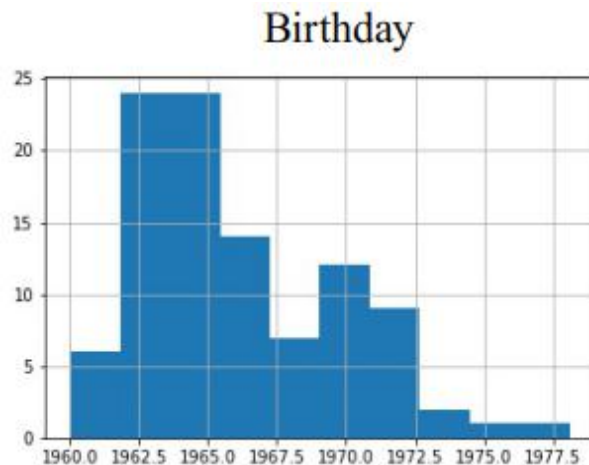


threshold 50



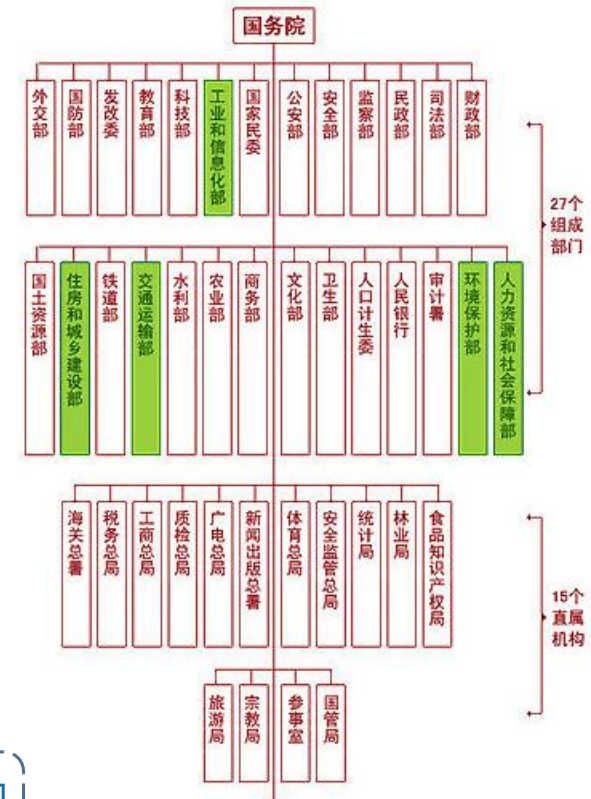
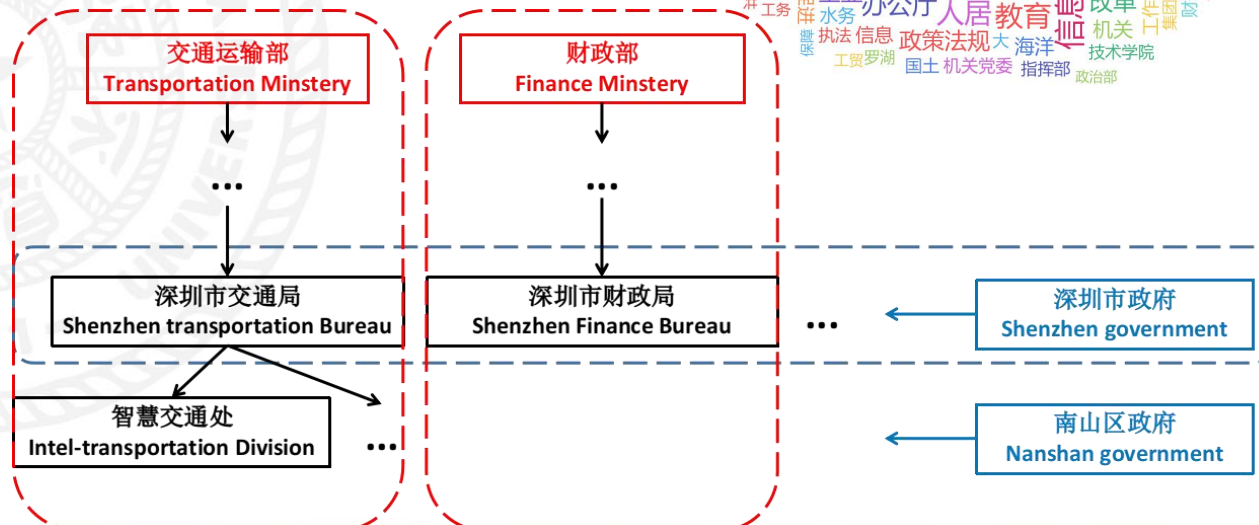
# Spatial-Temporal Graph: time step

- Choose smaller time step to observe the evolution clearly
- Another observation here is the distribution of our dataset is concentrate on 50-60 years old senior officers



# Spatial-Temporal Graph: Labeling

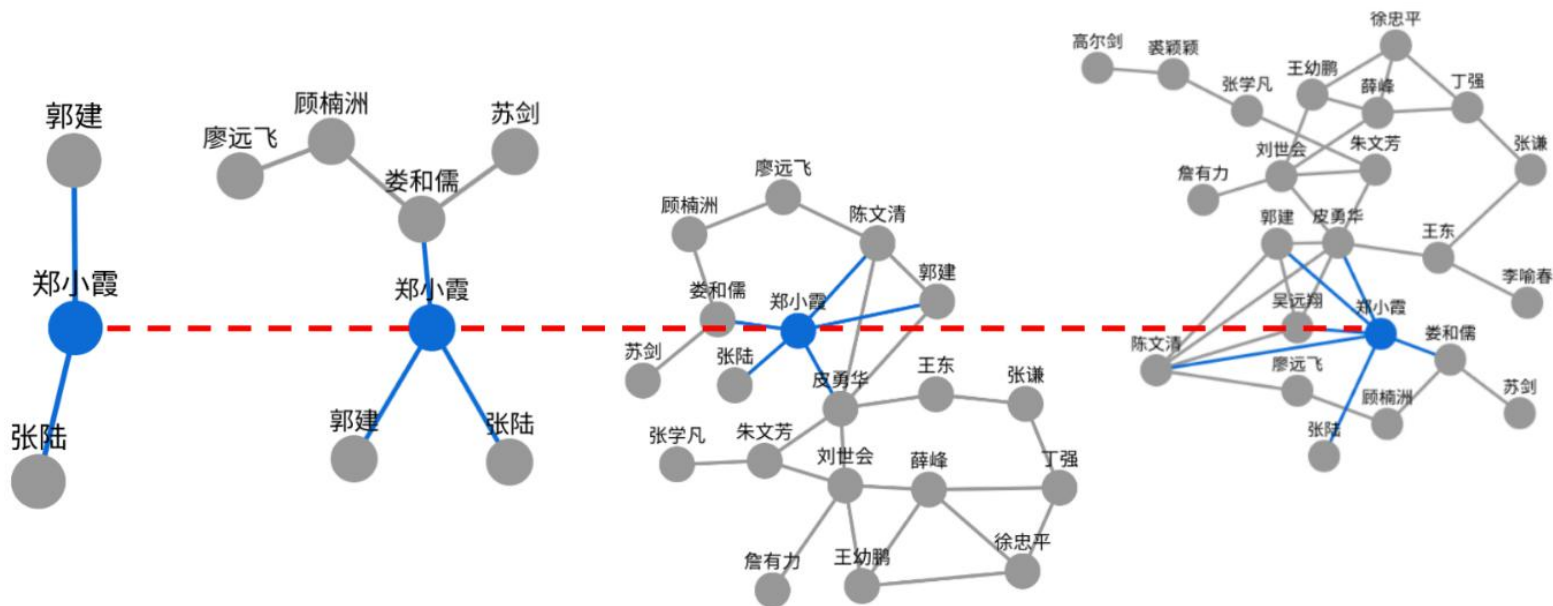
- Institution have Large diversity
- Share limited **purview of authority**  
(职能范围)vertically





# Spatial-Temporal Graph: Visualize

- Visualization in temporal domain shows the evolving characteristic in data



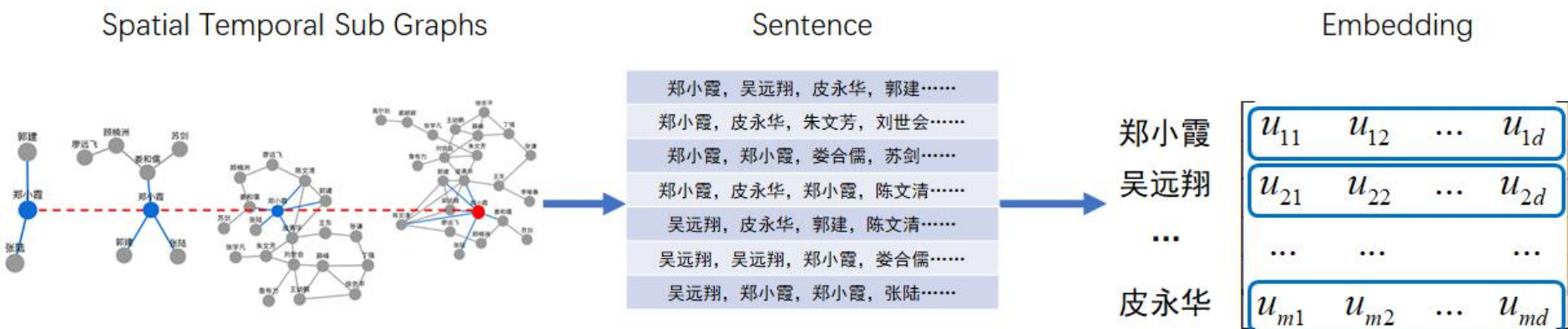


TBSI

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

# Data Analysis: STWalk

- A structure embedding method
- Two parts:
  - Random walk on the spatial-temporal network
  - SkipGram network to learn feature vector (NLP)



Supriya Pandhre, Himangi Mittal, Manish Gupta, and Vineeth N Balasubramanian. Stwalk: learning trajectory representations in temporal graphs. In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, pages 210–219. ACM, 2018.

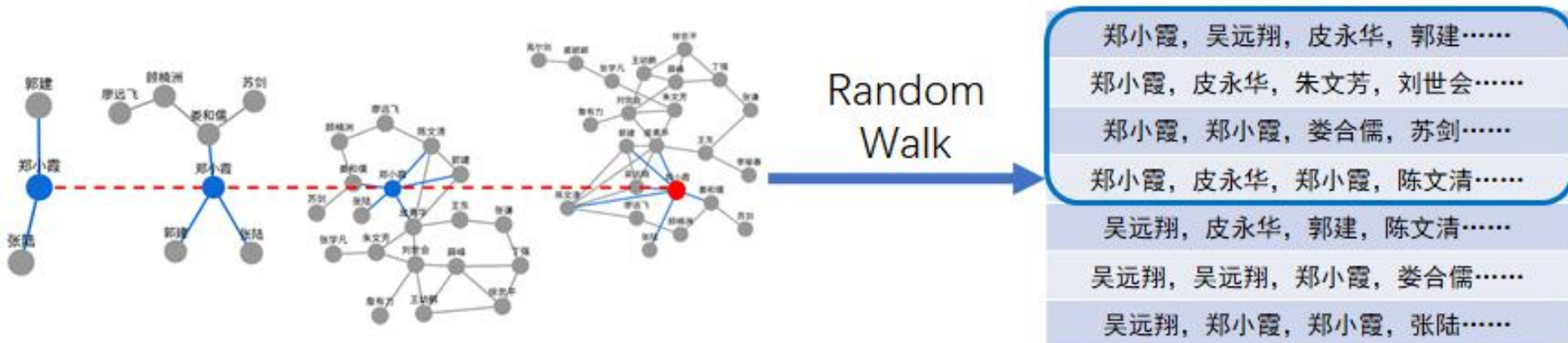


# Data Analysis: STWalk

- Random Walk in both spatial and temporal domain.
  - improvement: weighted sampling
- record the trajectory as 'sentence'

Spatial Temporal Sub Graph

Sentence







# Data Analysis: STWalk

- word embedding by SkipGram:
  - Maximizing the co-occurrence probability

$$\Pr(\phi(n_i) | \phi(u)) = \frac{\exp(\phi(n_i)\phi(u))}{\sum_{v \in N_u} \exp(\phi(v)\phi(u))}$$

Sentence

Embedding

郑小霞, 吴远翔, 皮永华, 郭建.....  
郑小霞, 皮永华, 朱文芳, 刘世会.....  
郑小霞, 郑小霞, 姜合儒, 苏剑.....  
郑小霞, 皮永华, 郑小霞, 陈文清.....  
吴远翔, 皮永华, 郭建, 陈文清.....  
吴远翔, 吴远翔, 郑小霞, 姜合儒.....  
吴远翔, 郑小霞, 郑小霞, 张陆.....

Word  
Embedding

郑小霞  
吴远翔  
...  
皮永华

|          |          |     |          |
|----------|----------|-----|----------|
| $u_{11}$ | $u_{12}$ | ... | $u_{1d}$ |
| $u_{21}$ | $u_{22}$ | ... | $u_{2d}$ |
| ...      | ...      | ... | ...      |
| $u_{m1}$ | $u_{m2}$ | ... | $u_{md}$ |



# Data Analysis: STWalk

---

## Algorithm 2 : STWalk

---

**Input:** Given node  $n \in [1, N]$ ,

time step  $t \in [1, T]$ ,

length of random walk:  $\mathcal{L}$ ,

temporal window size:  $\tau$ ,

vocabulary window size:  $\mathcal{W}_v$ ,

Set of graphs  $\mathcal{G}:\{G_t, G_{t-1}, \dots, G_{t-\tau}\}$ , size of embedding:  $d$ ,

number of restarts (starts at same node):  $\rho$

**Output:** Updated  $n^{th}$  row of matrix of  $\Phi_t$  of size:  $N \times d$

- 1 Initialize  $\Phi_t[n] \in \mathbb{R}^d$
  - 2 **for**  $i=0$  to  $\rho$  **do**
  - 3      $\text{new\_graph} = \text{createSpaceTimeGraph}(\mathcal{G}, \tau, n)$
  - 4      $\text{spaceTimeWalk} = \text{randomWalk}(\text{new\_graph}, n, \mathcal{L});$
  - 5      $\Phi_t[n] = \text{SkipGram}(\Phi_t, \text{spaceTimeWalk}, \mathcal{W}_v);$
  - 6 **end**
-



TBSI

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

# Data Analysis: Classification

- Multilabel classification problem
  - One-hot encoding to encode labels
- One-Vs-The-Rest classification problem
  - Multiple Binary SVM classifier

| Time | 任继光                     | 刘伟                                 |
|------|-------------------------|------------------------------------|
| t    | 市律师协会,市公证协会<br>(司法, 司法) | 龙岗区政府副区长, 发展和改革委员会综合处<br>(机关, 发改委) |
| t-1  | 深圳市司法局<br>(司法)          | 人民政府办公厅财经处, 综合处<br>(机关, 财政)        |
| t-2  | 宝安区司法局<br>(司法)          | 复旦大学<br>(高等院校)                     |

0  
1 机关  
0  
0  
0  
.  
.  
1 发改委

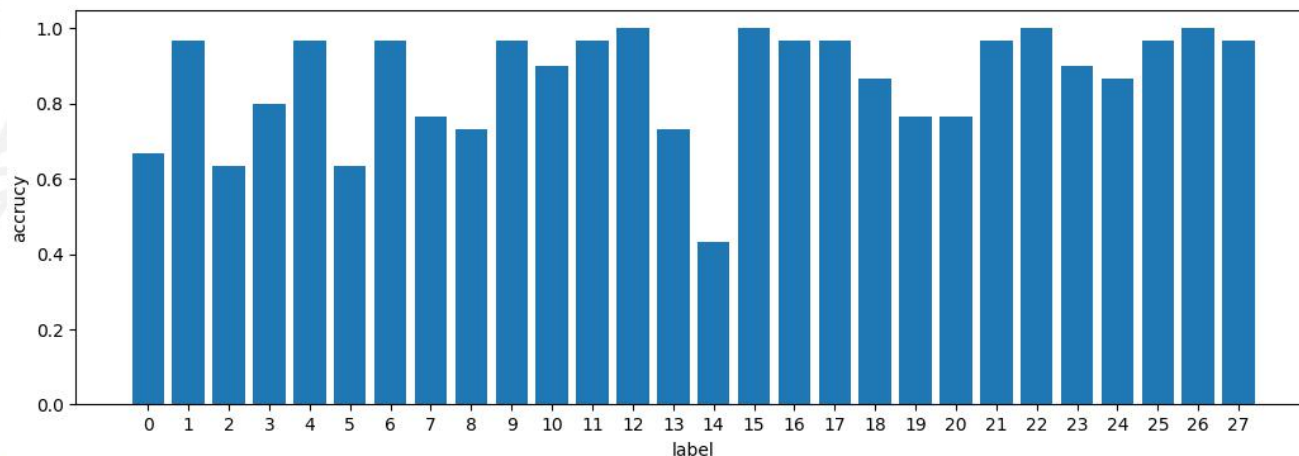


TBSI

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



# Result: Multilabel classification

- In average, the accuracy of classification is 86.19%
- The accuracy of correctly predict at least one label for a node is 73.33%.
- Label accuracy





# Data Analysis: Classification

- Interdisciplinary or expert problem:
  - Only one label  Expert
  - More than one label  Interdisciplinary

|      | Expert                  | Interdisciplinary                  |
|------|-------------------------|------------------------------------|
| Time | 任继光                     | 刘伟                                 |
| t    | 市律师协会,市公证协会<br>(司法, 司法) | 龙岗区政府副区长, 发展和改革委员会综合处<br>(机关, 发改委) |
| t-1  | 深圳市司法局<br>(司法)          | 人民政府办公厅财经处, 综合处<br>(机关, 财政)        |
| t-2  | 宝安区司法局<br>(司法)          | 复旦大学<br>(高等院校)                     |

- In average, the accuracy of classification is 86.67%\*



# Future work

- Extend the scope of dataset to over the country
- Promotion analysis and prediction
- Better embedding with node features





TBSI

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

# Team Contribution

陈彦锋

Yanfeng Chen

Data Crawling and Parsing

毛适

Shi Mao

Spatial-Temporal Graph  
Construction and Visualization

王远

Yuan Wang

Data Analysis





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

# THANK YOU!

