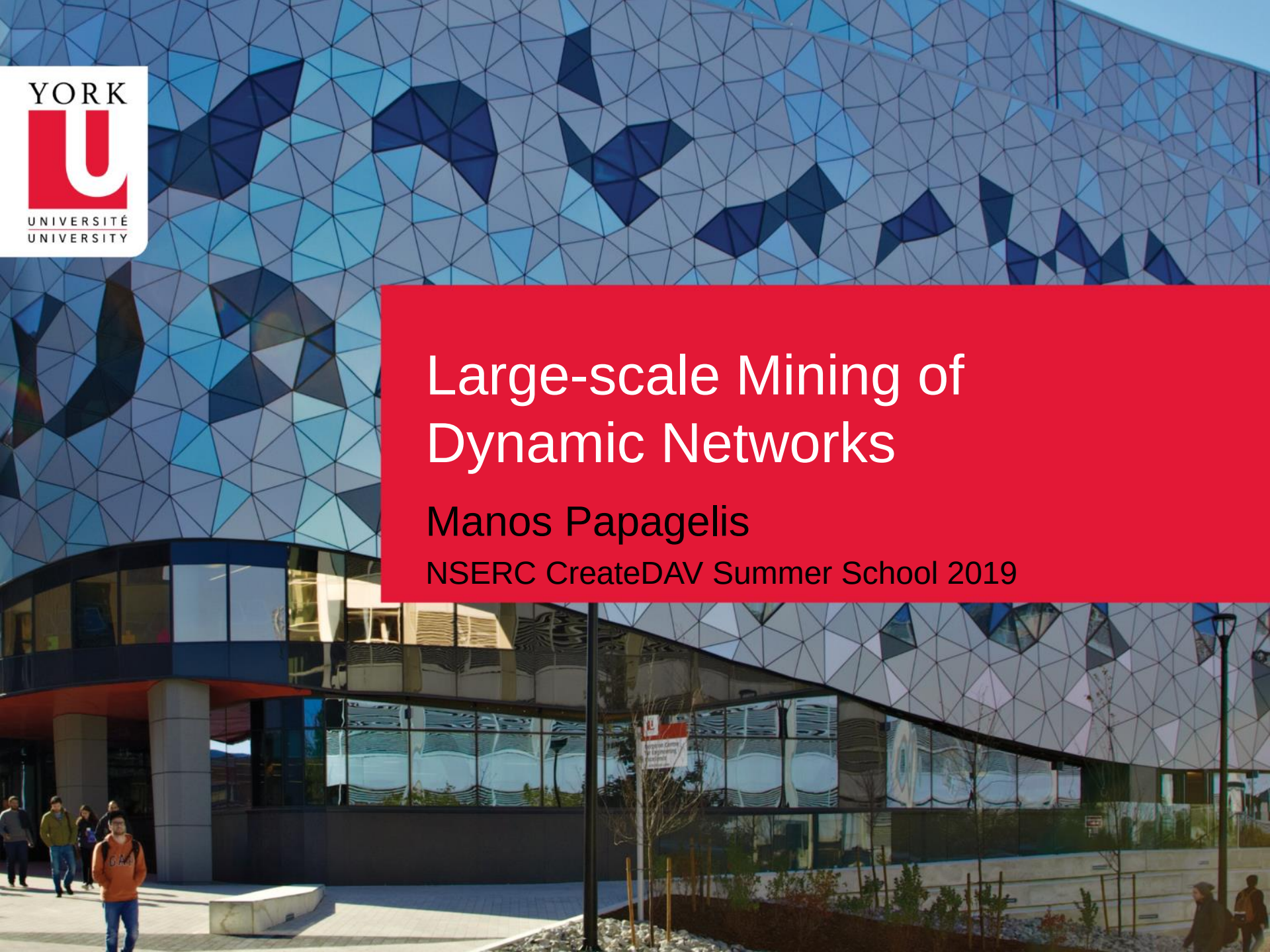


Large-scale Mining of Dynamic Networks

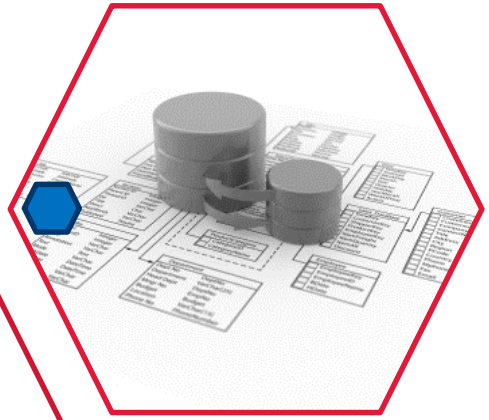
Manos Papagelis

NSERC CreateDAV Summer School 2019

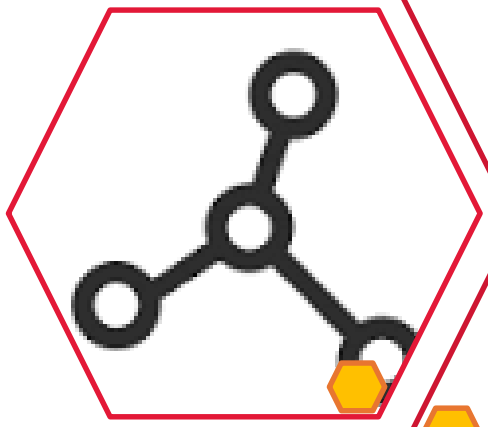
Background


UNIVERSITY OF TORONTO
YORK UNIVERSITÉ UNIVERSITY
University of Crete
FORTH
Institute of Computer Science
Berkeley
UNIVERSITY OF CALIFORNIA
YAHOO!
LABS

Research



Big Data & Knowledge Discovery



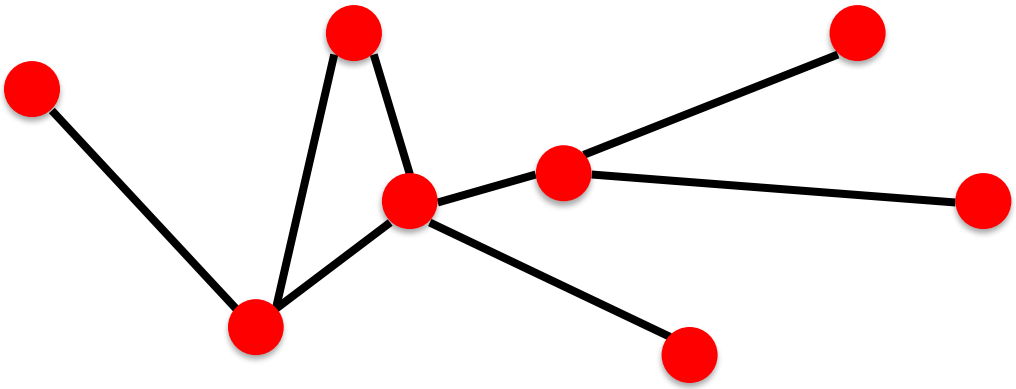
Data Mining

Graph Mining

Today

what is a network or a graph?

Network Components



- **Network** (or **Graph**) $G\ (N, E)$
  - **Objects**: nodes (vertices) N
  - **Relationships:** links (edges) E

Built on the mathematics of **graph theory**

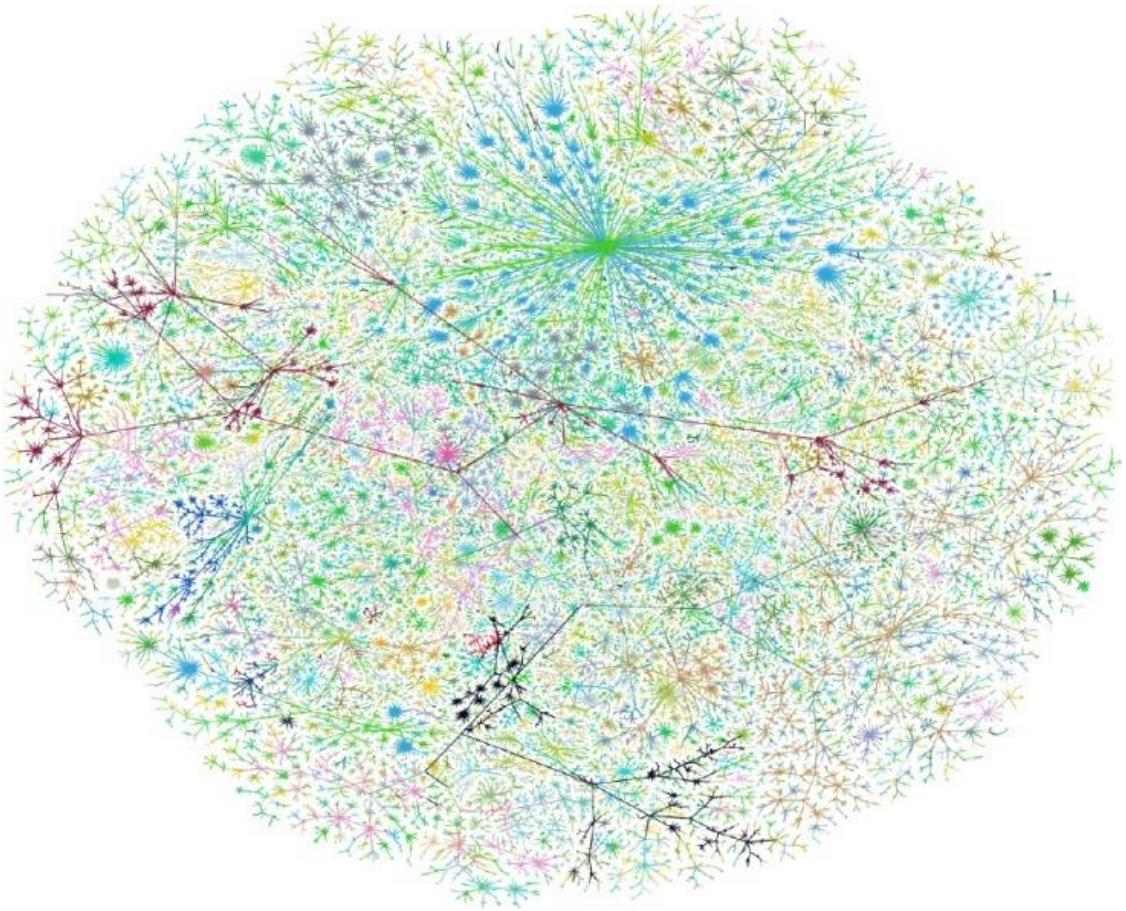
networks are ubiquitous

Networks: Social

Facebook social graph

4-degrees of separation [Backstrom-Boldi-Rosa-Ugander-Vigna, 2011]

Networks: Communication

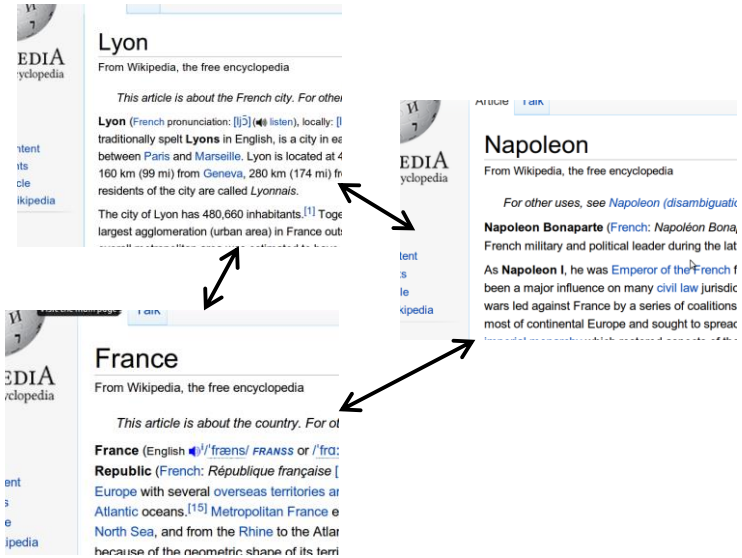


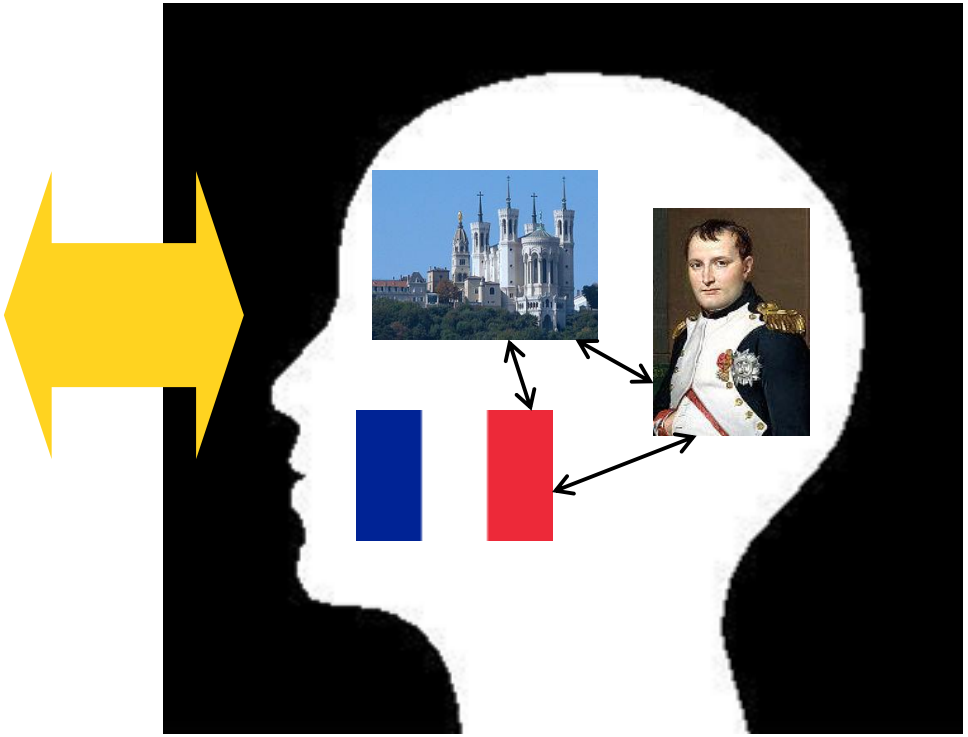
Graph of the Internet (Autonomous Systems)

Power-law degrees [Faloutsos-Faloutsos-Faloutsos, 1999]

Robustness [Doyle-Willinger, 2005]

Networks: Knowledge Graph



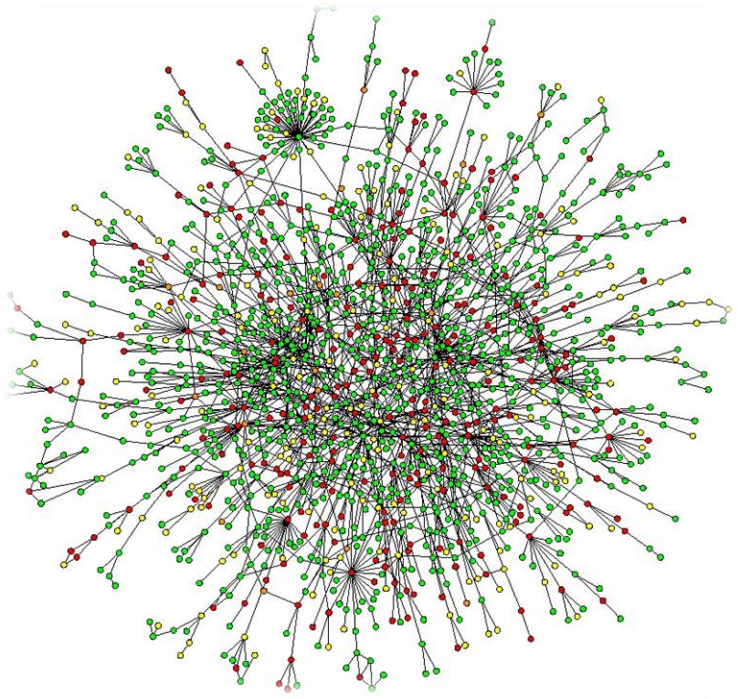


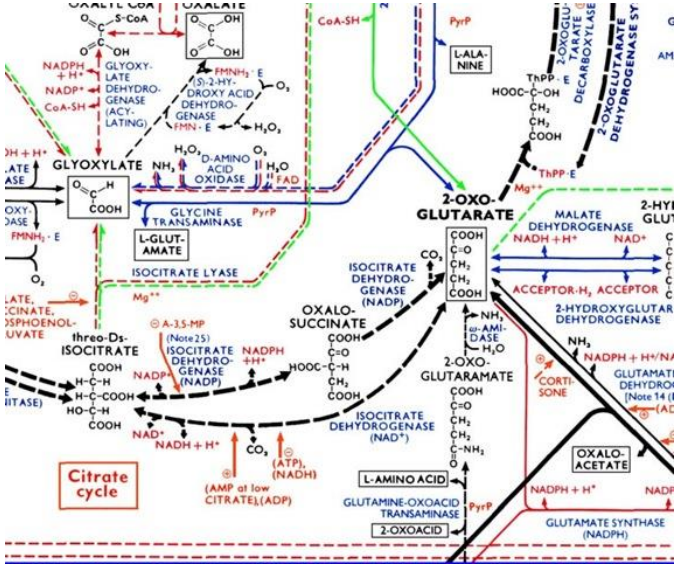
Understand how humans navigate Wikipedia

Get an idea of how people connect concepts

[West-Leskovec, 2012]

Networks: Biology





Protein-Protein Interaction Networks:

Nodes: Proteins

Edges: ‘physical’ interactions

Metabolic networks:

Nodes: Metabolites and enzymes

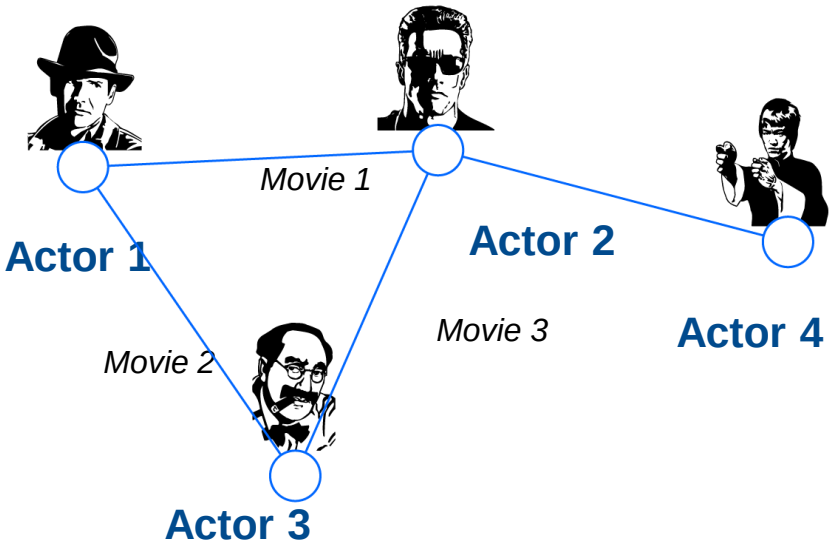
Edges: Chemical reactions

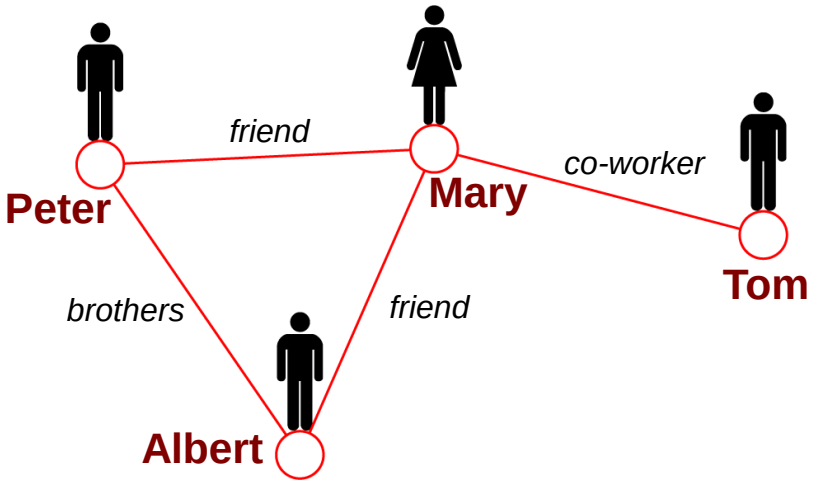
Networks: Brain

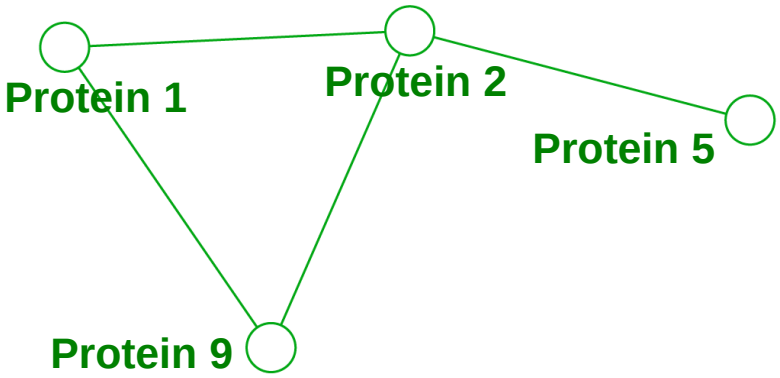
Human brain has between
10-100 billion neurons
[Sporns, 2011]

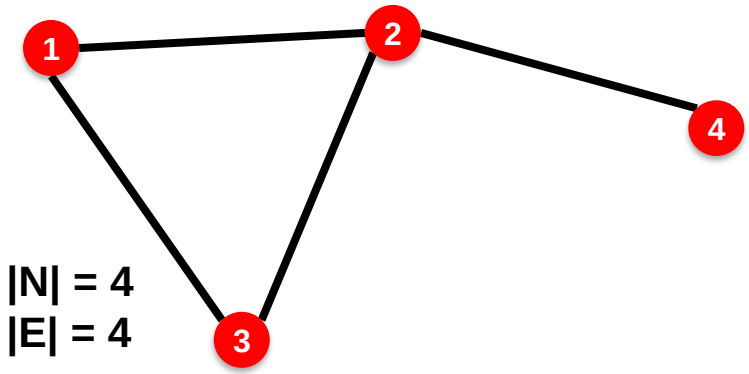
why should we care about networks?

Networks: Common Language









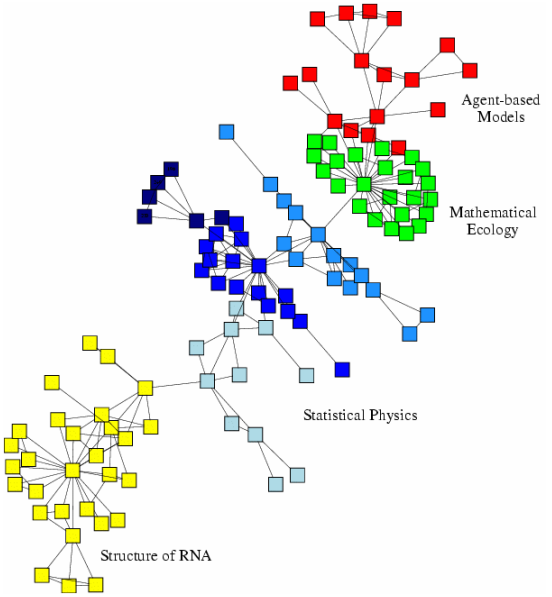
Network Analysis

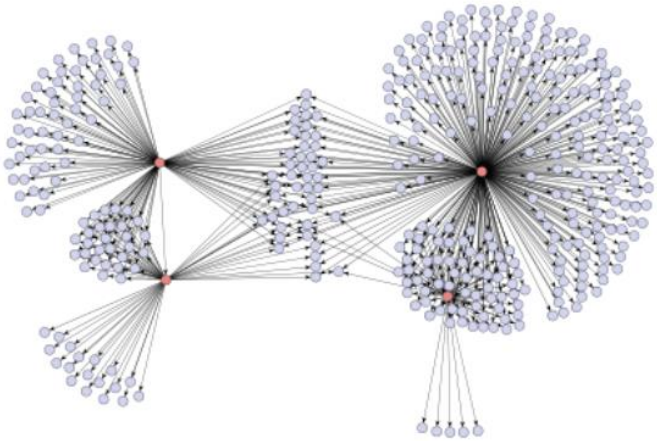
network analysis helps to reveal the *underlying dynamics* of these systems, not easily observable before

what do we study in networks?

Networks: Structure & Process

- **Structure and evolution**
  - What is the structure of a network?
  - Why and how did it become to have such structure?
- **Processes and dynamics**
  - Networks provide “skeleton” for spreading of information, behavior, diseases





how do we reason about networks?

Reasoning About Networks

- **Empirical studies/properties**

  Study network data to find organizational principles

- **Mathematical models**

  Probabilistic, graph theory

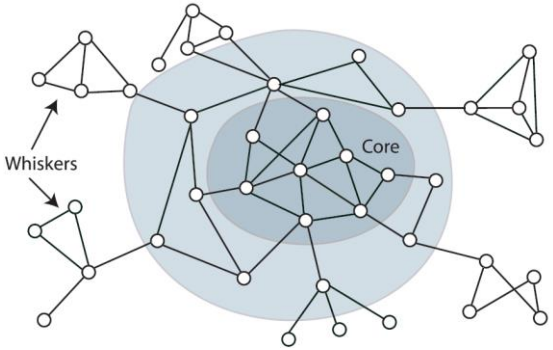
- **Algorithms**

  Methods for analyzing graphs, solving graph-related problems

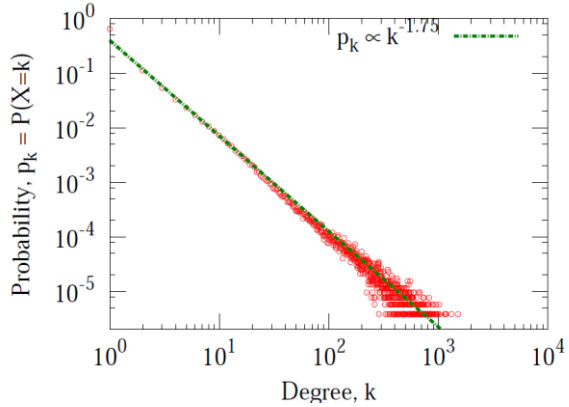
Properties

- **Six degrees of separ.**

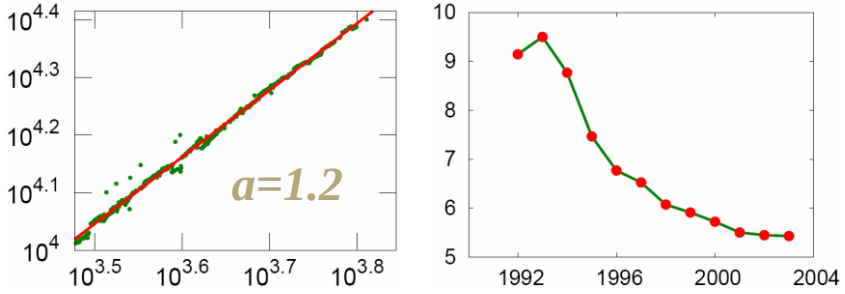
- **Strength of weak ties**



- **Power-law degrees**

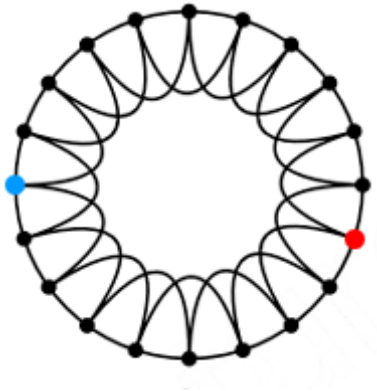


- **Densif. power law, Shrinking diameter**

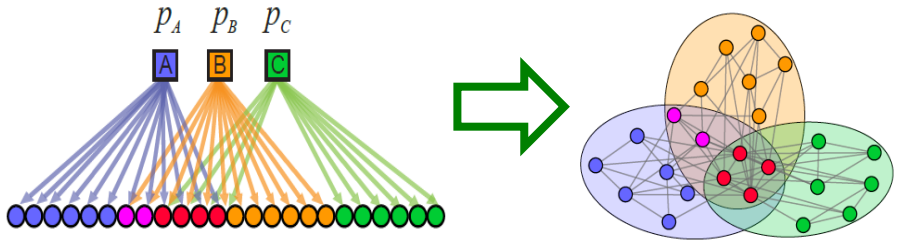


Models

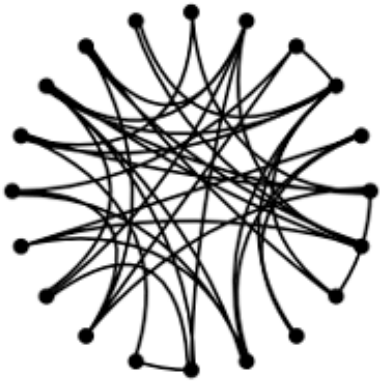
- **Erdös-Renyi model**



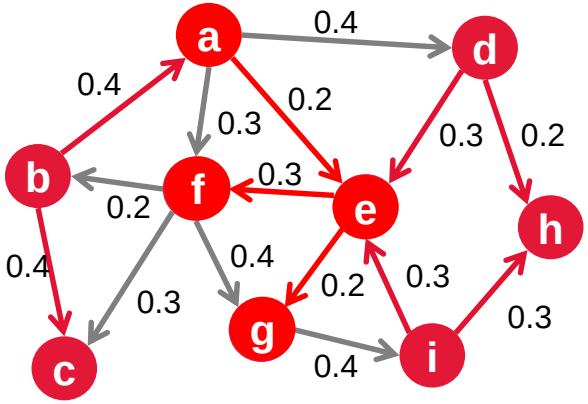
- **Community model**



- **Small-world model**

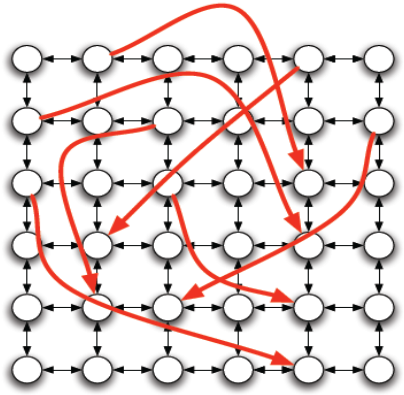


- **Cascade model**

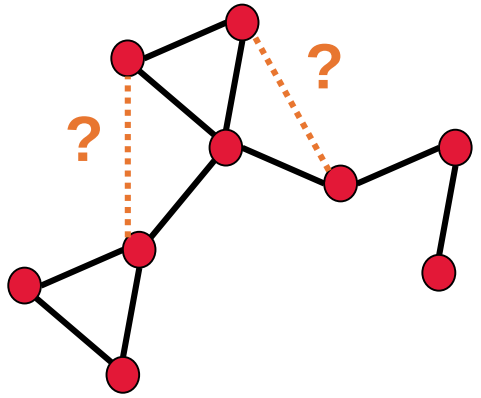


Algorithms

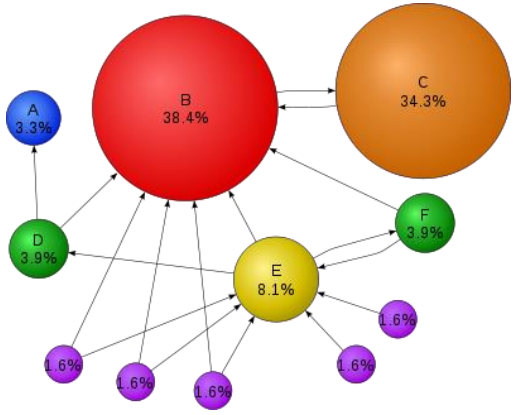
- **Decentralized search**



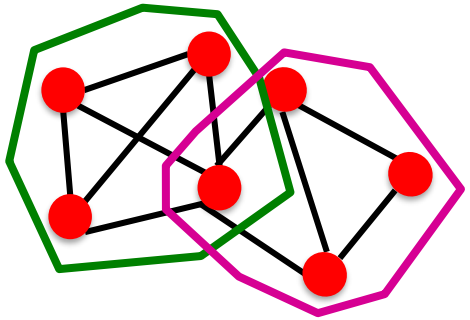
- **Link prediction**



- **Link analysis**



- **Community detection**



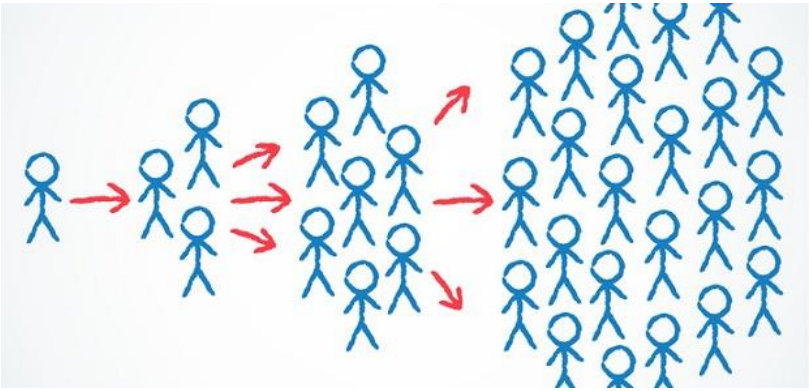
Map of Superpowers

Properties	Models	Algorithms
Small diameter, Edge clustering	Small-world model, Erdös-Renyi model	Decentralized search
Scale-free	Preferential attachment, Copying model	PageRank, Hubs and authorities
Strength of weak ties, Core-periphery	Community-affiliation Graph Model	Community detection: Girvan-Newman, Modularity
Densification power law, Shrinking diameters	Microscopic model of evolving networks	Link prediction, Supervised random walks
Information virality, reproductive number	Independent cascade model, Game theoretic model, SIR	Influence maximization, Outbreak detection, LIM

Applying Our Superpowers

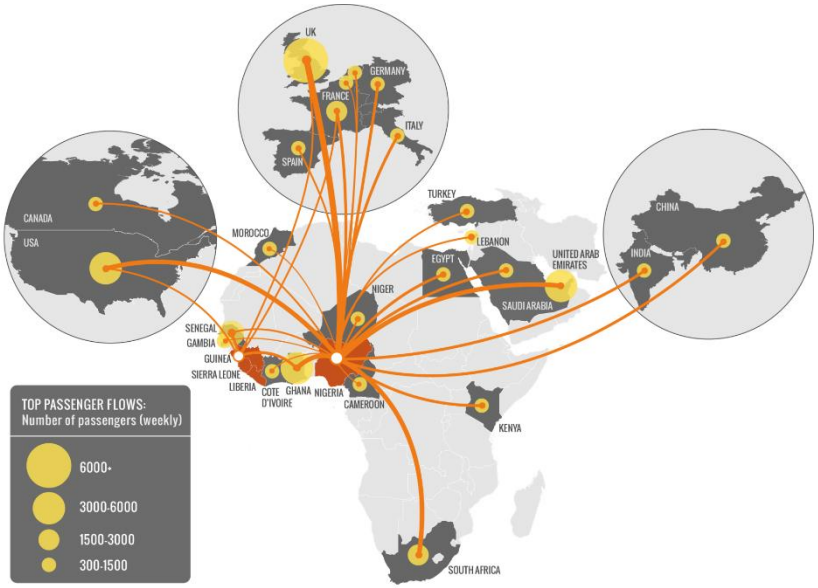
- **Social media analytics**

- **Viral marketing**

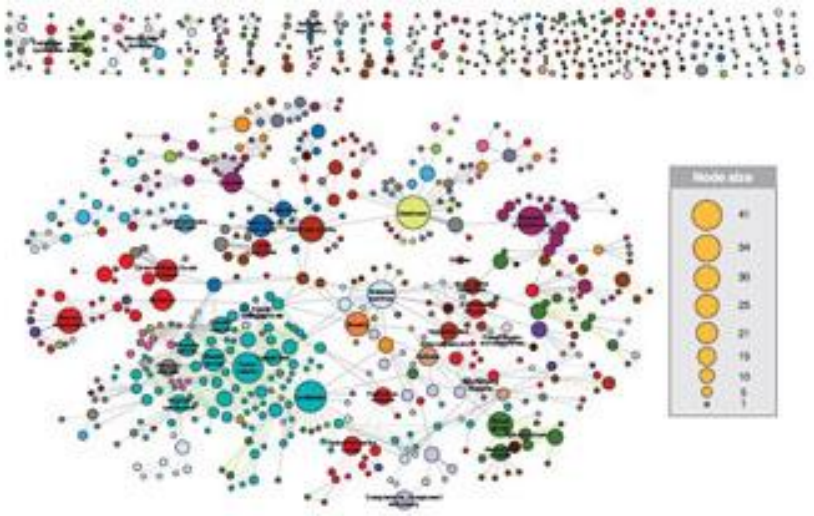


Applying Our Superpowers

- **Predicting epidemics: Ebola**



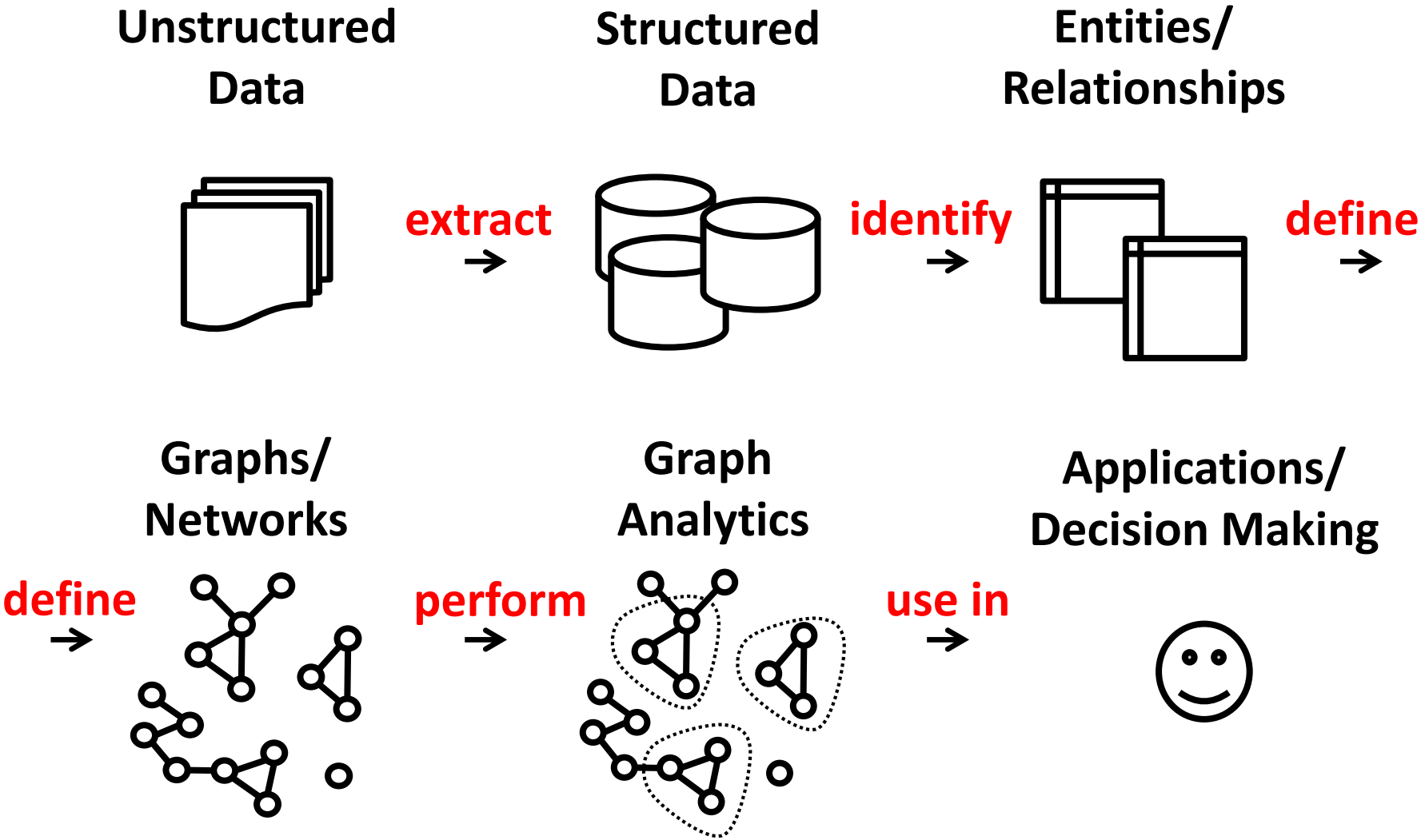
- **Drug design**



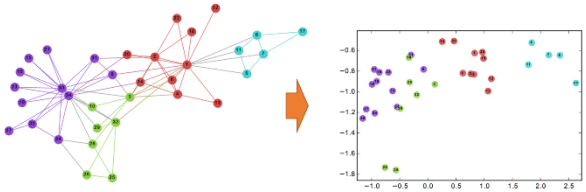
Graph Mining
Abstract Methodology

YORK
UNIVERSITÉ
UNIVERSITY

Research Methodology

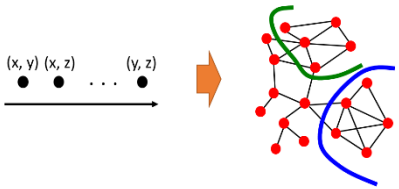


Current Research Focus



A. Network Representation Learning

B. Trajectory Network Mining



C. Streaming & Dynamic Graphs

D. Social Media Mining & Analysis

E. City Science / Urban Informatics / IoT

F. Natural Language Processing

Trajectory Network Mining

Problem 1

Group Pattern Discovery of Pedestrian Trajectories

Problem 2

Mining of Node Importance in Trajectory Networks

Group Pattern Discovery of Pedestrian Trajectories

Joint work with Sawas Abdullah et al.

Pedestrian trajectories

what is a group?

many definitions, many algorithms

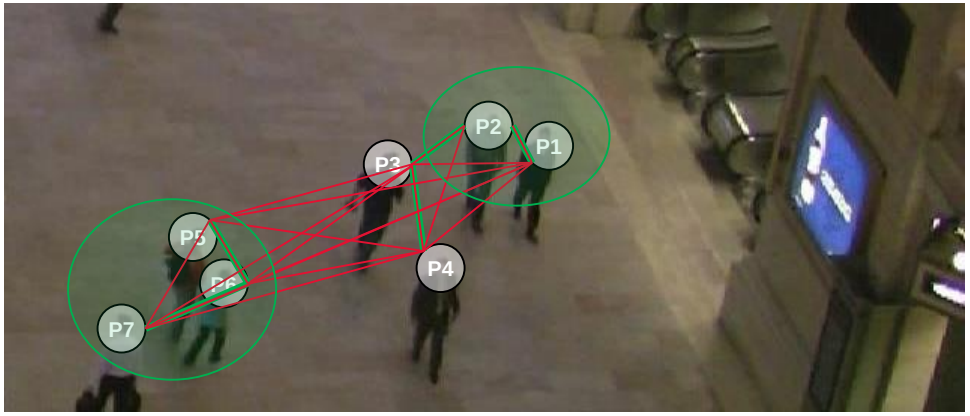
e.g., ***flock***, ***convoy***, ***evolving-clusters***, ***gathering-pattern***, … [ACM TIST Tutorial 2015]

Finding pedestrian groups

Local Grouping

Intuitive method

Spatial-only



proximity threshold θ

key idea

find **pairs** of pedestrians **x**, **y** where **distance(x, y) < θ**

expand **pairs** to discover **groups**

Local grouping

expand the key idea
to include the
time dimension

Global groups vs. Time-window groups

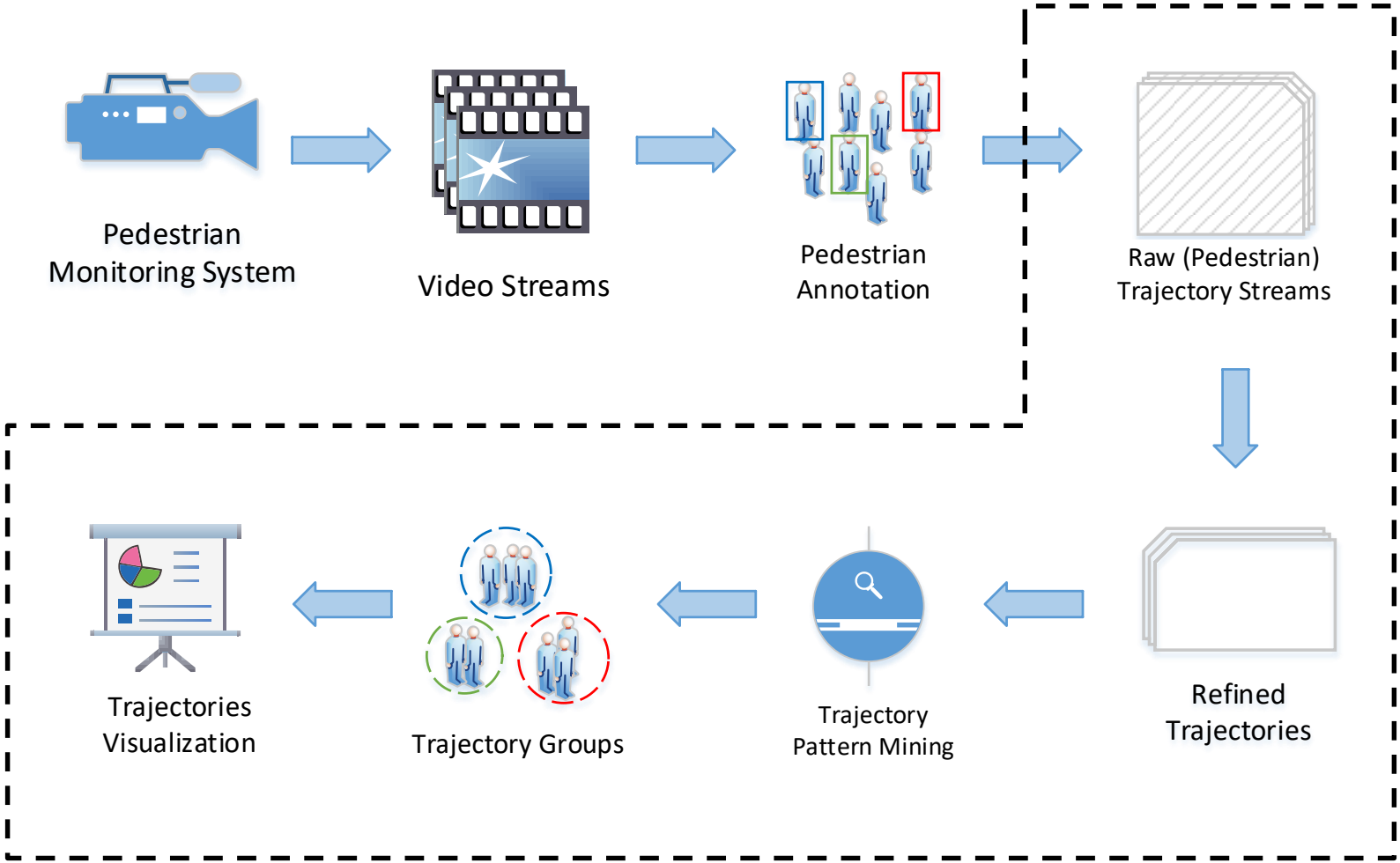
global grouping

time-window grouping

Trajectolizer

Demo

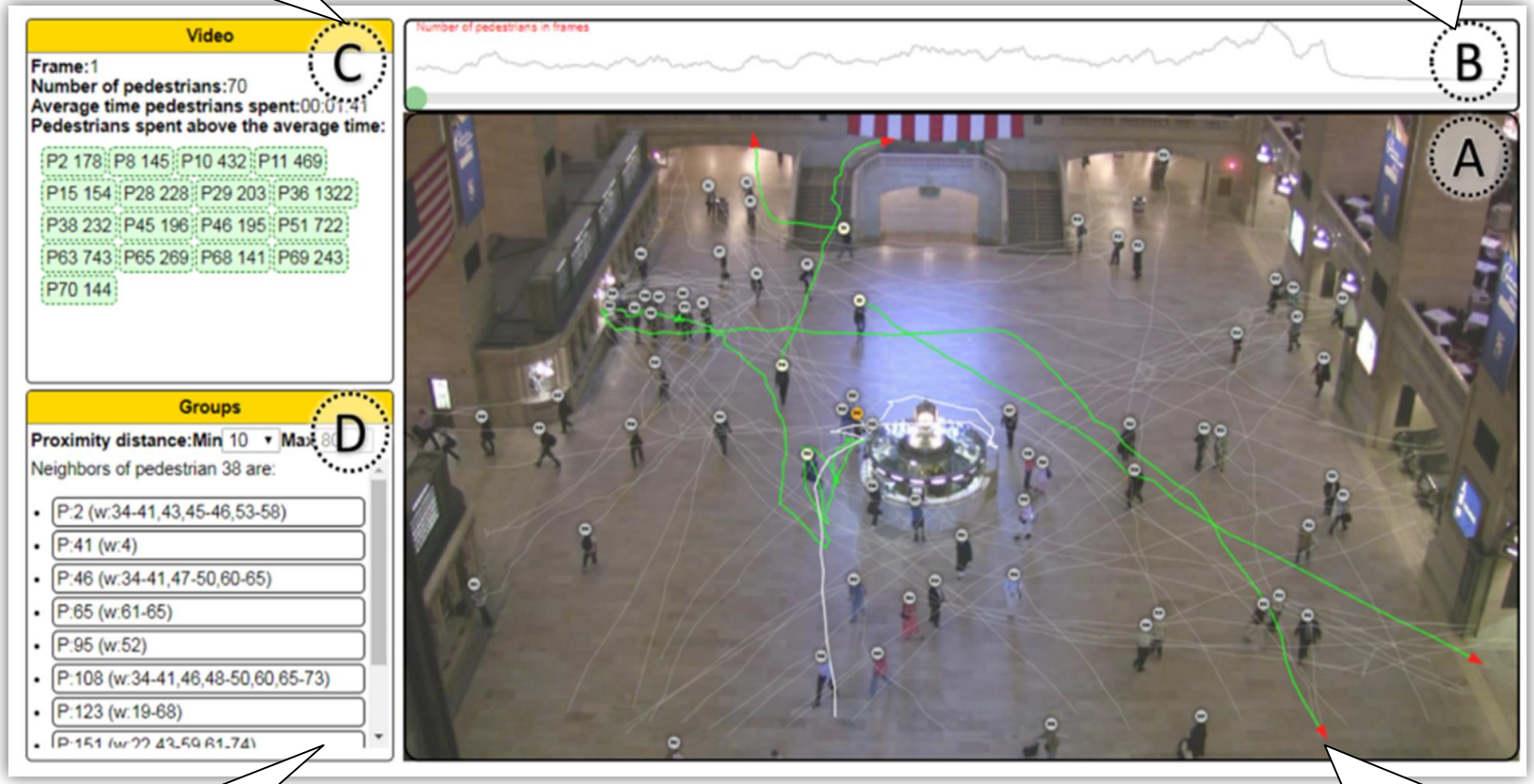
Trajectolizer: System Overview



Trajectolizer: Interactive Demo

descriptive statistics about the current frame

timeline slider area to navigate video frames



grouping analysis

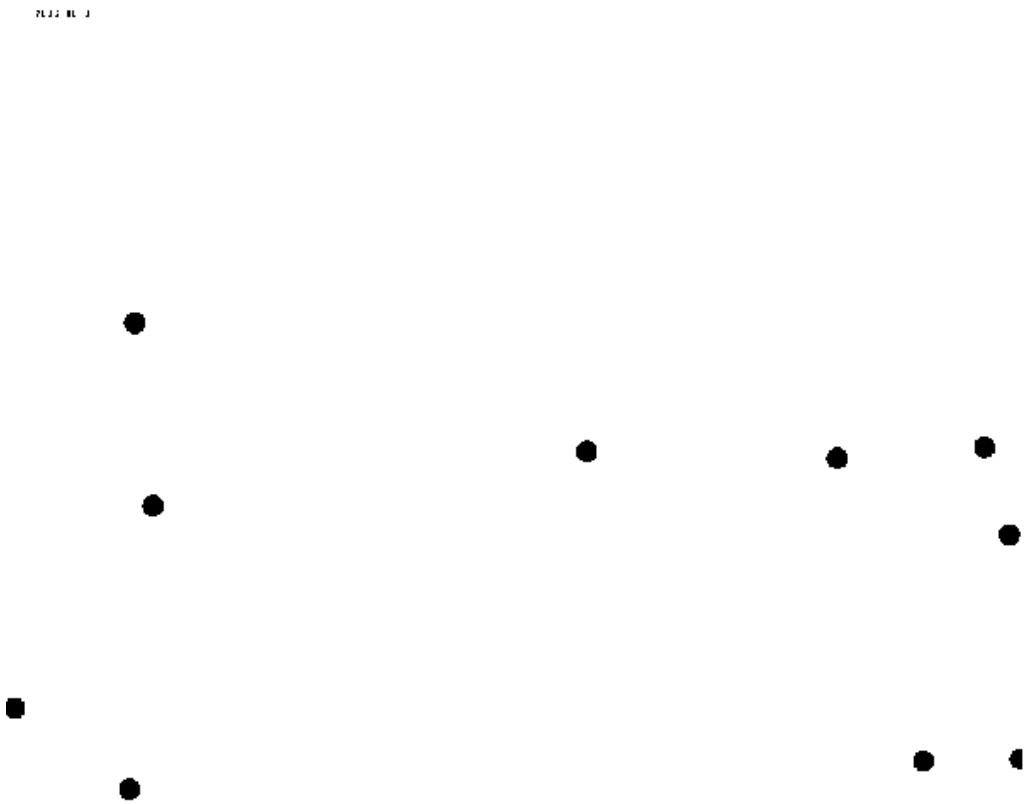
current frame with pedestrian IDs and trajectories

Live Demo

Node Importance in Trajectory Networks

Joint work with Tilemachos Pechlivanoglou

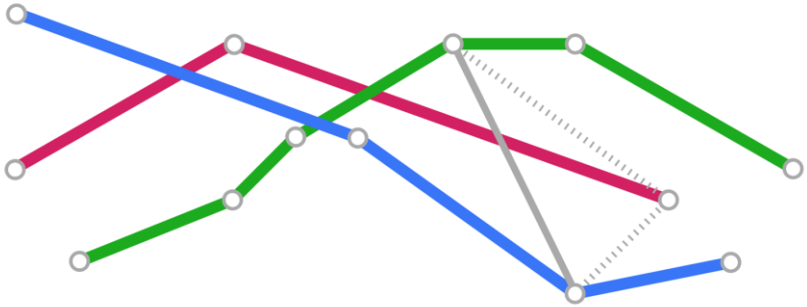
Trajectories of moving objects



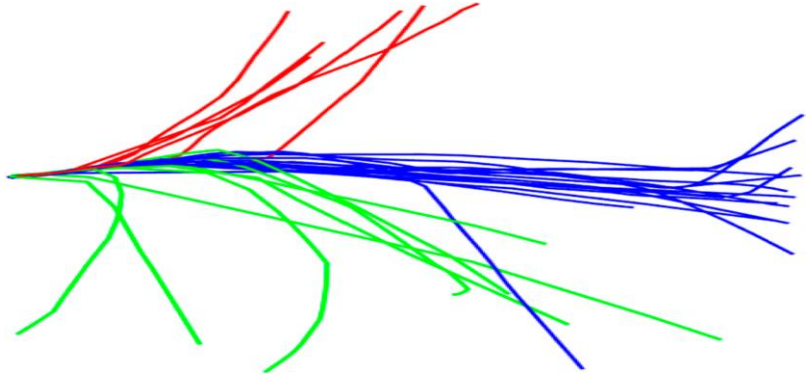
every moving object, forms a **trajectory** – in **2D** it is a sequence of (**x**, **y**, **t**)

there are trajectories of moving **cars**, **people**, **birds**, …

Trajectory data mining



trajectory similarity

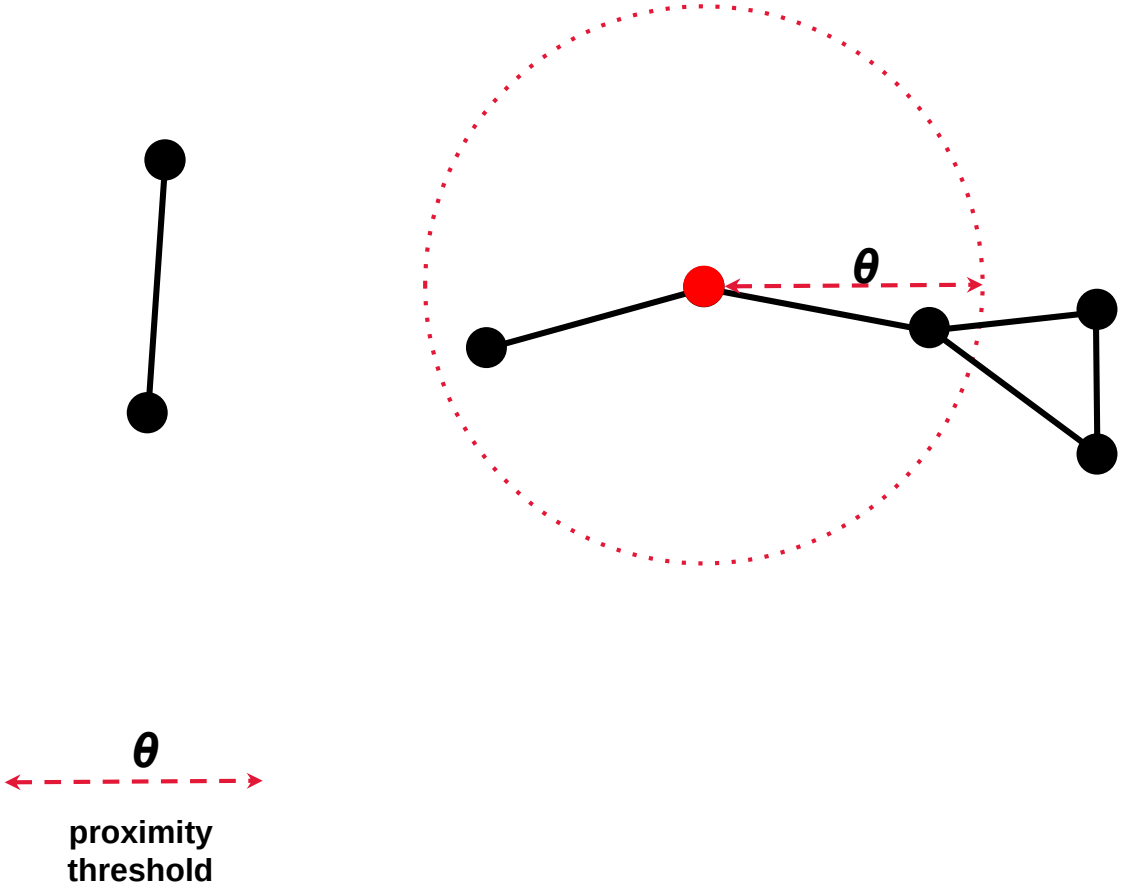


trajectory clustering

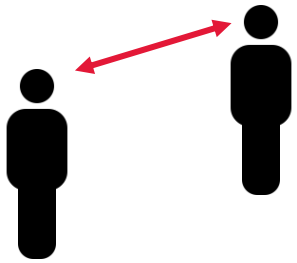
trajectory anomaly detection
trajectory pattern mining
trajectory classification
...more

we care about network analysis of moving objects

Proximity networks



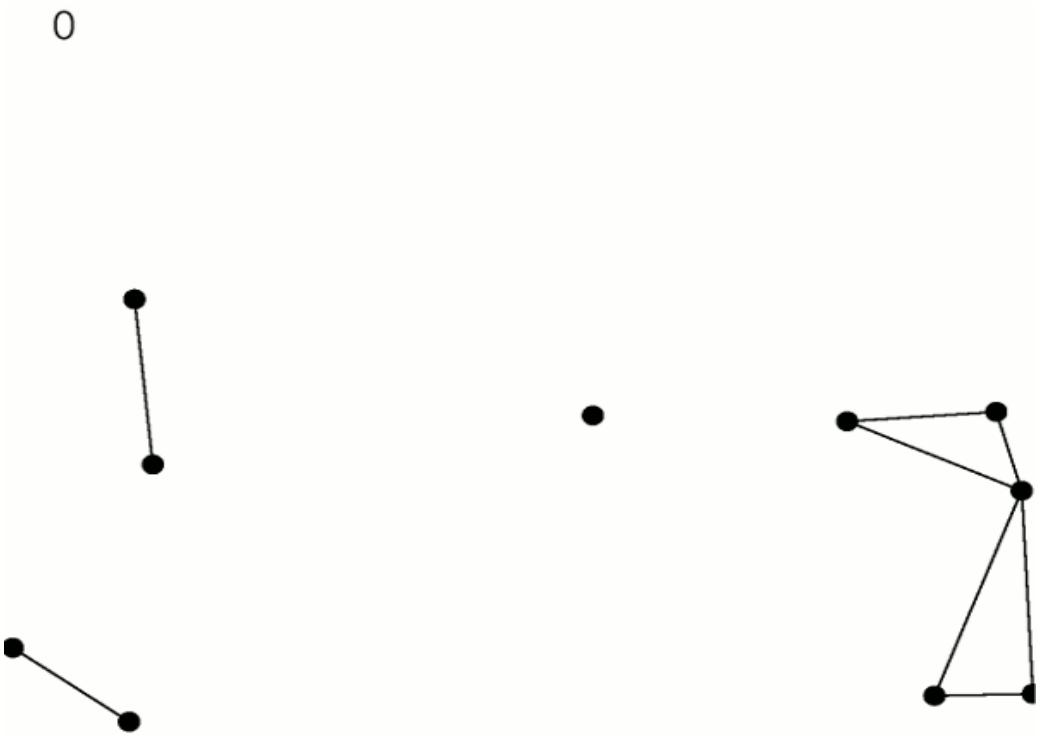
Distance can represent



line of sight

wifi / bluetooth
signal range

Trajectory networks



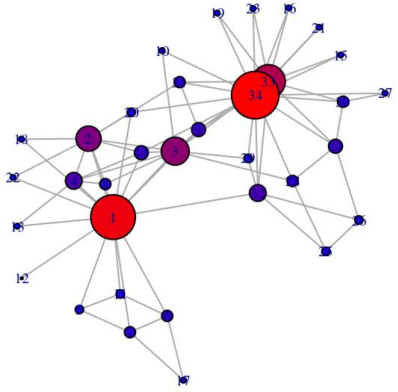
The Problem

Input: logs of trajectories (**x**, **y**, **t**) in time period [0, T]

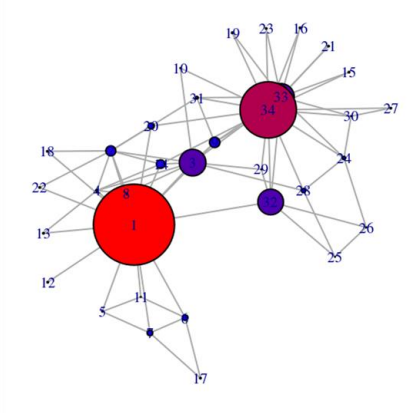
Output: node importance metrics

Node Importance

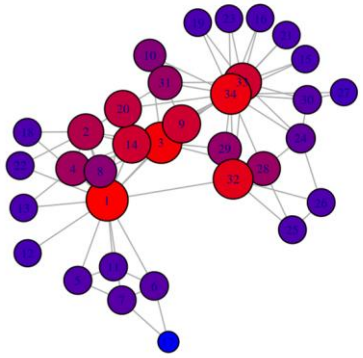
Node importance in static networks



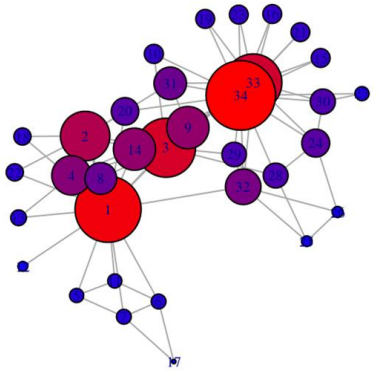
Degree centrality



Betweenness centrality

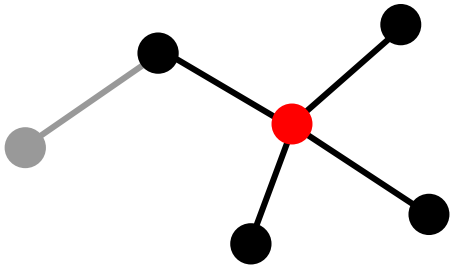


Closeness centrality

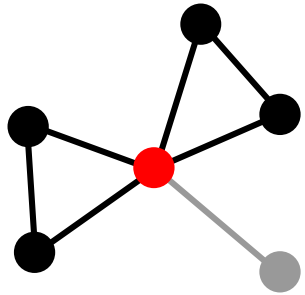


Eigenvector centrality

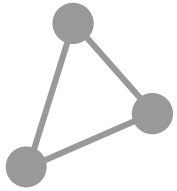
Node importance in TNs

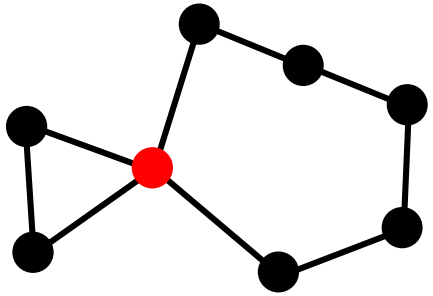


node degree **over time**



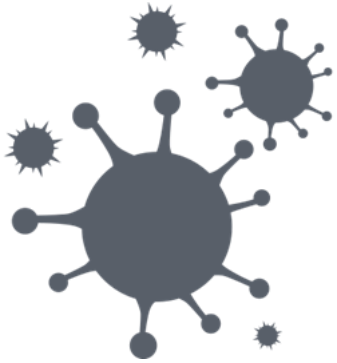
triangles **over time**





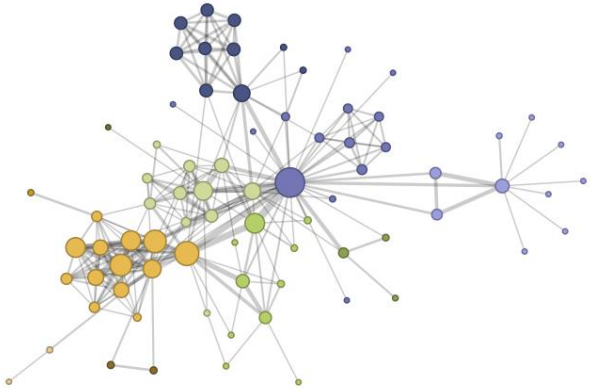
connected components **over time**
(connectedness)

Applications



infection spreading

security in autonomous vehicles

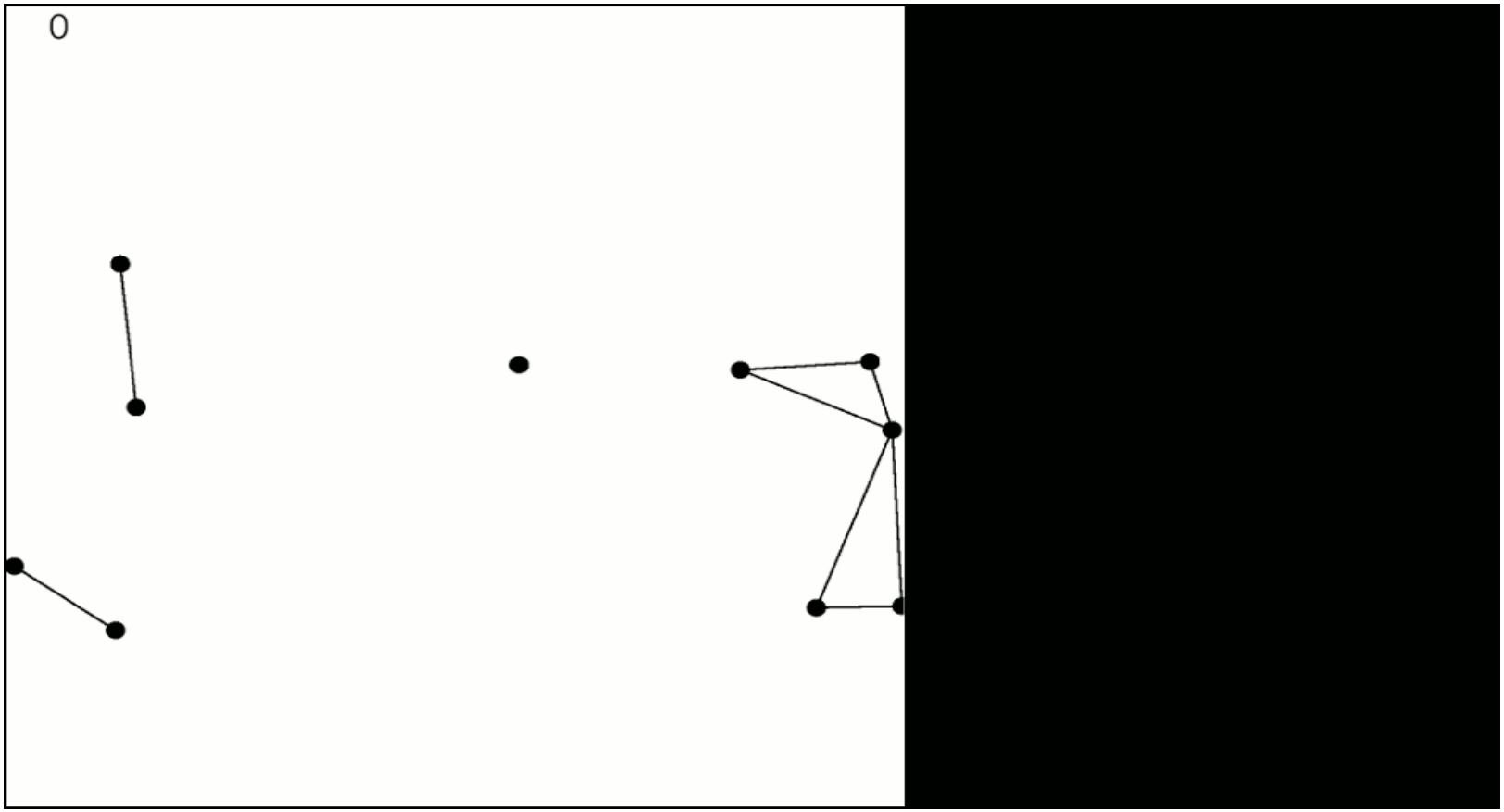


rich dynamic network analytics

Evaluation of Node Importance in Trajectory Networks

YORK UNIVERSITÉ UNIVERSITY

Naive approach



For **every** discrete time unit **t**:

1. obtain static **snapshot** of the proximity network
2. run **static** node importance **algorithms** on snapshot

Aggregate results at the end

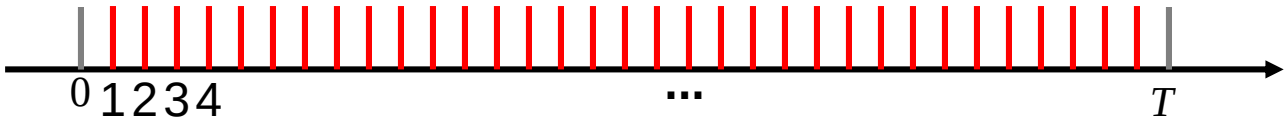
Streaming approach

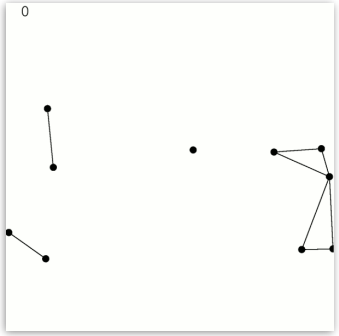
Similar to naive, but:

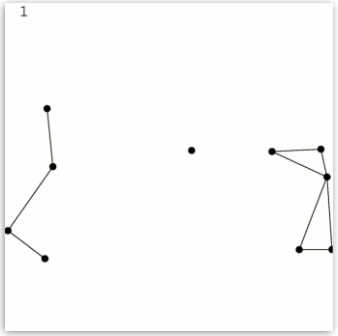
- **no final aggregation**
- results calculated **incrementally** at every step

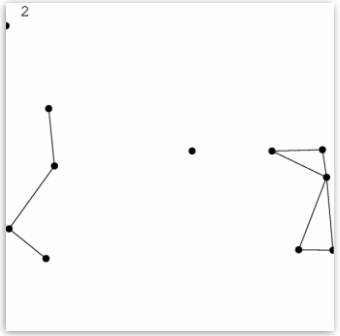
Still **every time** unit

Every discrete time unit







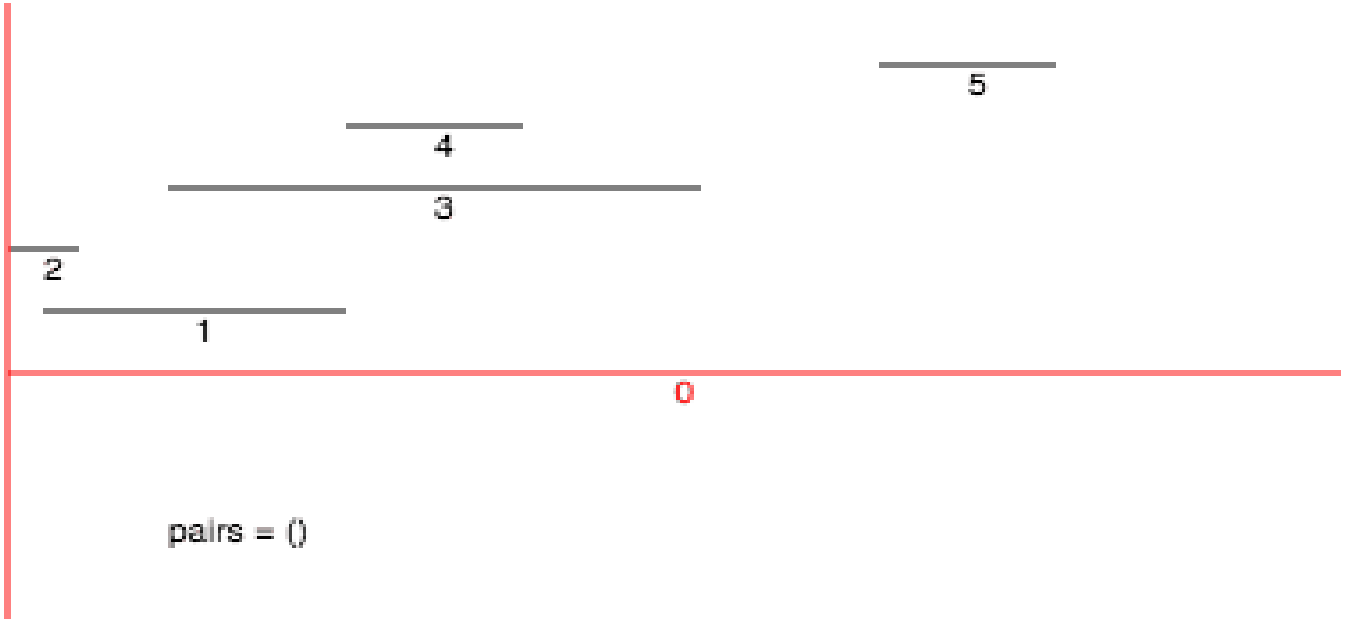


...

Sweep Line Over Trajectories (SLOT)

Sweep line algorithm

A computational geometry algorithm that given **line segments** computes line segment **overlaps**



Efficient **one pass** algorithm that only processes line segments at the **beginning** and **ending** points

SLOT: **S**weep **L**ine **O**ver **T**rajectories

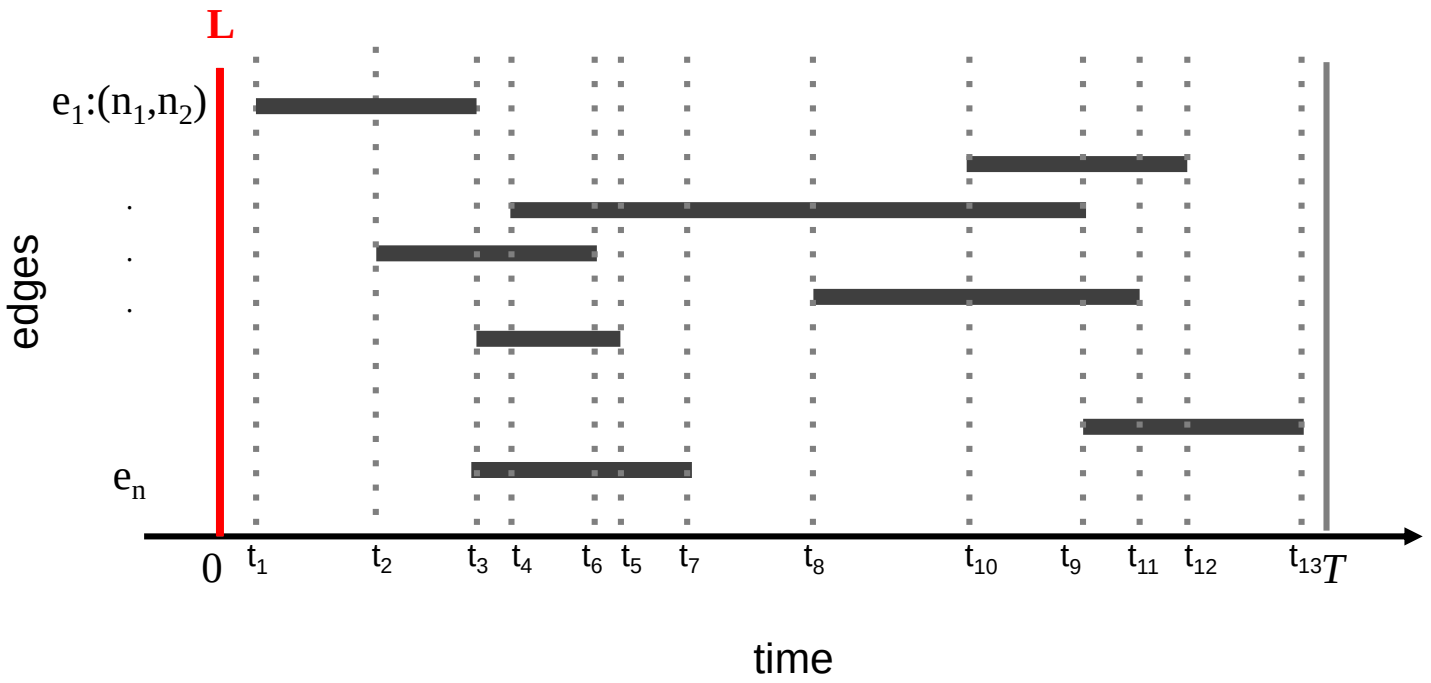
(**algorithm sketch**)

represent TN **edges** as **time intervals**

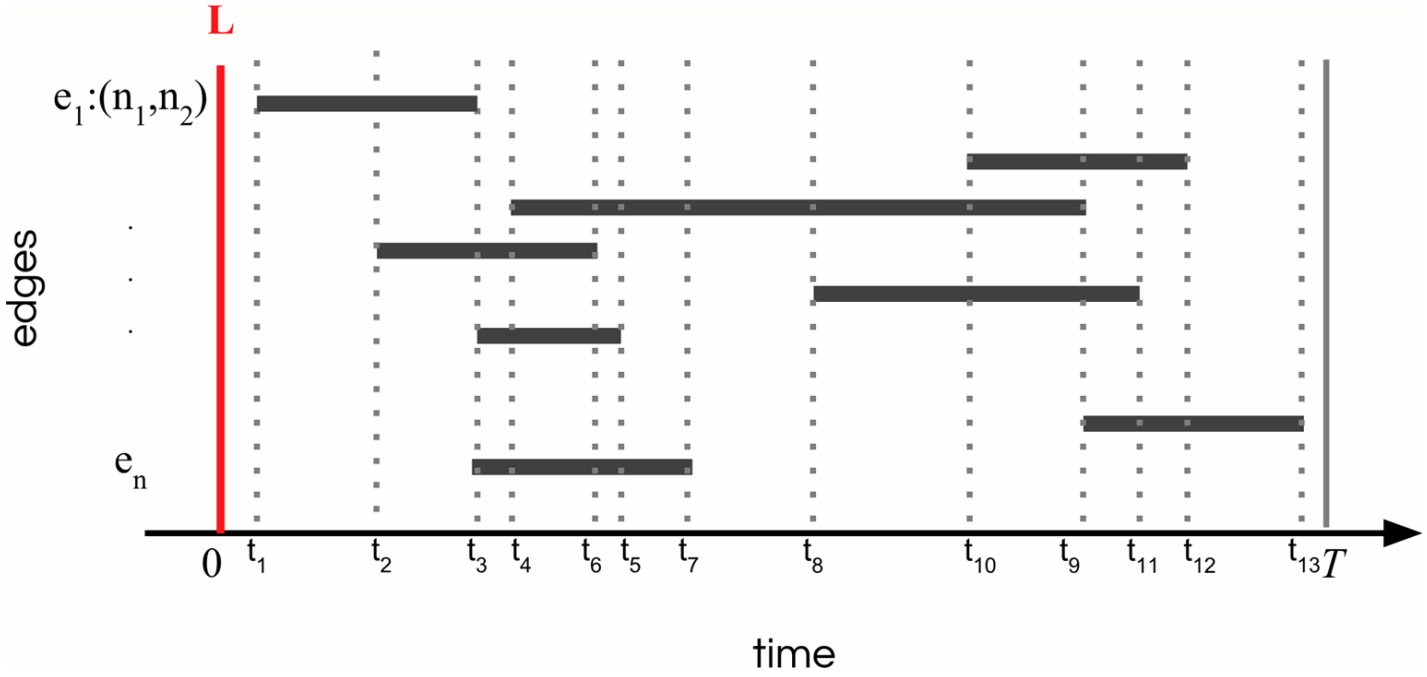
apply **variation** of sweep line algorithm

simultaneously compute *node degree*, *triangle membership*, *connected components* in **one pass**

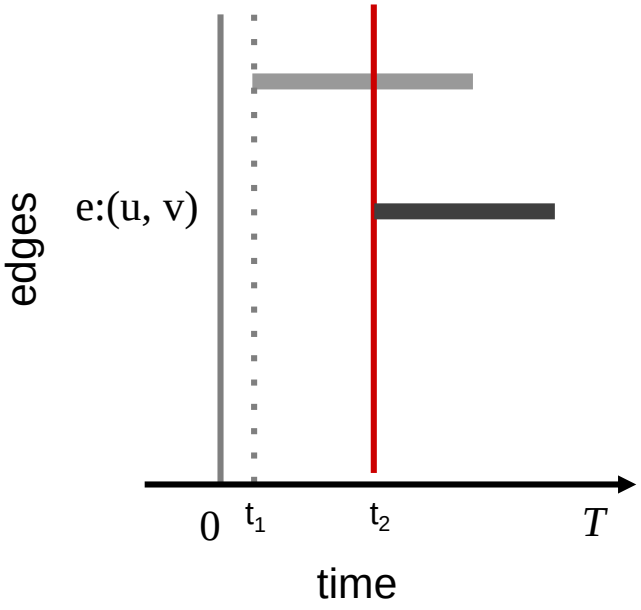
Represent edges as time intervals

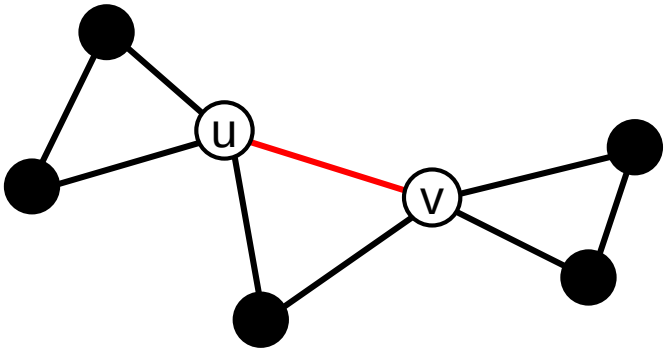


SLOT: **S**weep **L**ine **O**ver **T**rajectories



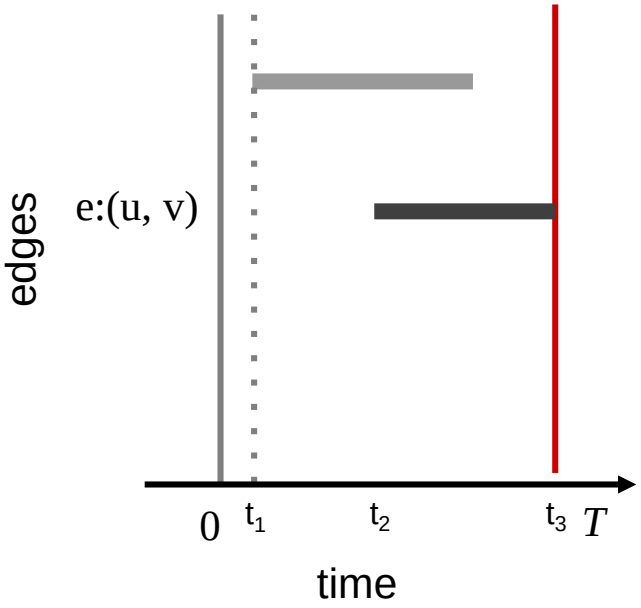
At every edge **start**

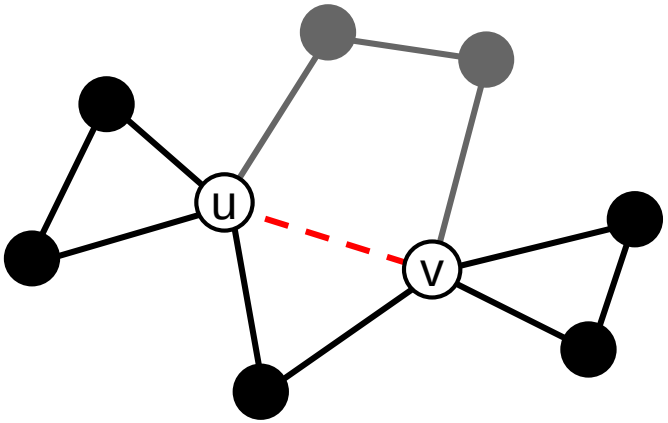




- **node degree**
  - nodes u, v now connected
  - increment u, v node degrees
- **triangle membership**
  - did a triangle just form?
  - look for u, v common neighbors
  - increment triangle (u, v, common)
- **connected components**
  - did two previously disconnected components connect?
  - compare old components of u, v
  - if no overlap, merge them

At every edge **stop**





- **node degree**
  - nodes u, v now disconnected
  - decrement u, v degree
- **triangle membership**
  - did a triangle just break?
  - look for u, v common neighbors
  - decrement triangle (u, v, common)
- **connected components**
  - did a conn. compon. separate?
  - BFS to see if u, v still connected
  - if not, split component to two

SLOT: At the end of the algorithm …

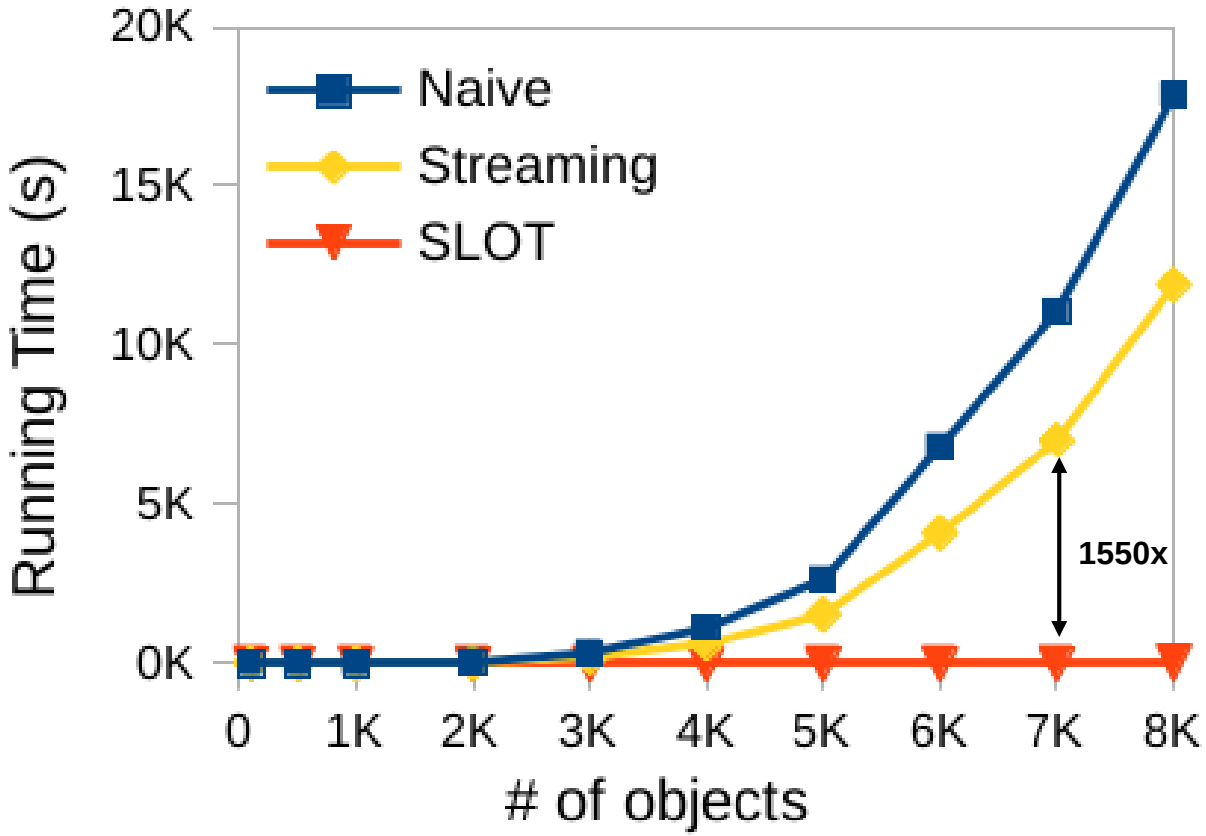
Rich Analytics

- **node degrees**: start/end time, duration
- **triangles**: start/end time, duration
- **connected components**: start/end time, duration

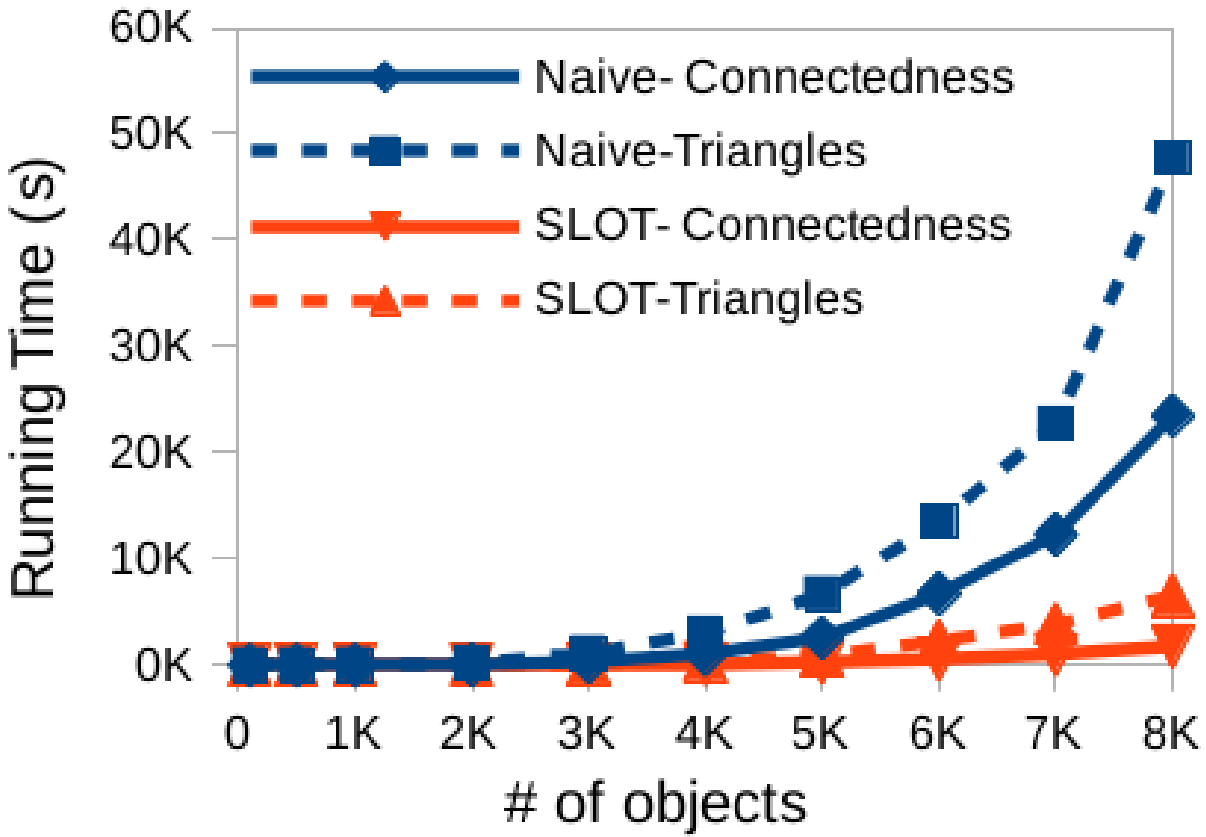
Exact results (not approximations)

Evaluation of SLOT

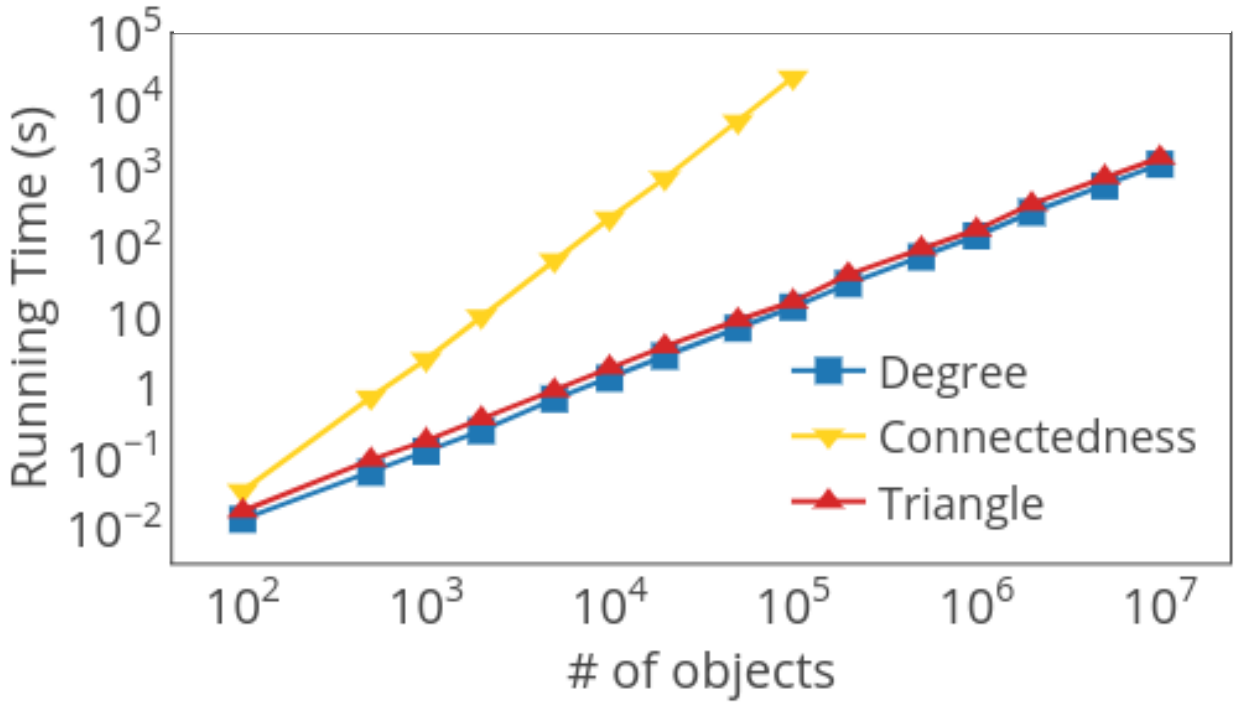
Node degree


20K
15K
10K
5K
0K
Running Time (s)
Naive
Streaming
SLOT
1550x
0
1K
2K
3K
4K
5K
6K
7K
8K
# of objects

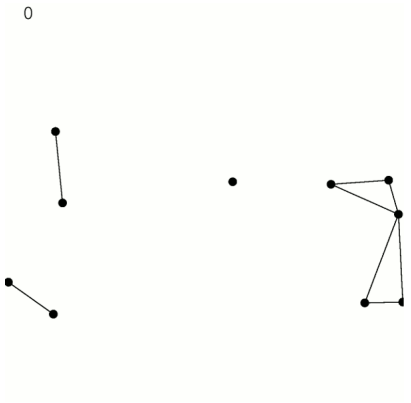
Triangle membership / connected components



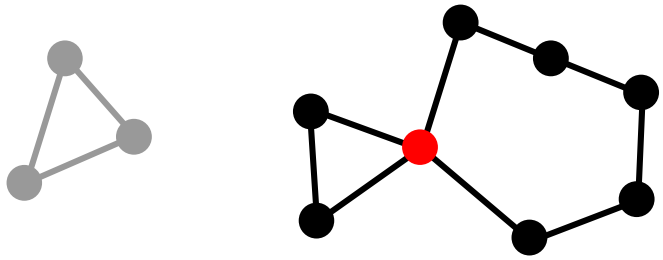
SLOT Scalability



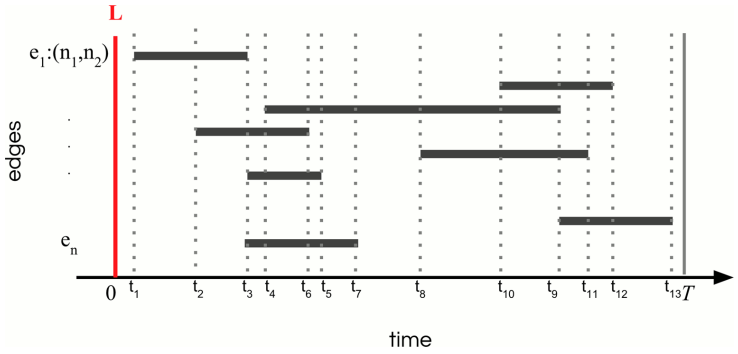
Takeaway



trajectory networks



network importance **over time**



SLOT algorithm

SLOT properties:

- fast
- exact
- scalable

Seagull migration trajectories

data from Wikelski et al. 2015

Thank you!

Questions?

References

[Geoinformatica 2019] **A Versatile Computational Framework for Group Pattern Mining of Pedestrian Trajectories**. Abdullah Sawas, Abdullah Abuolaim, Mahmoud Afifi, Manos Papagelis. GeoInformatica (Vol. X, No. X, 2019)

[IEEE Big Data 2018] **Fast and Accurate Mining of Node Importance in Trajectory Networks**. Tilemachos Pechlivanoglou and Manos Papagelis. (IEEE Big Data 2018)

[IEEE MDM 2018] **Tensor Methods for Group Pattern Discovery of Pedestrian Trajectories**. Abdullah Sawas, Abdullah Abuolaim, Mahmoud Afifi, Manos Papagelis. Proceedings of the 19th IEEE International Conference on Mobile Data Management (IEEE MDM 2018, **best paper award**)

[IEEE MDM 2018] **Trajectolizer: Interactive Analysis and Exploration of Trajectory Group Dynamics**. Abdullah Sawas, Abdullah Abuolaim, Mahmoud Afifi, Manos Papagelis. Proceedings of the 19th IEEE International Conference on Mobile Data Management (IEEE MDM 2018, demo)

Working with Us

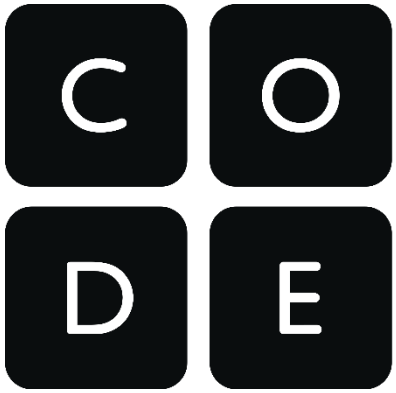
YORK
UNIVERSITÉ
UNIVERSITY

Data Mining Lab @ YorkU

- **Data Mining Lab**
  - http://dminer.eecs.yorku.ca/
  - **focus**: data mining / machine learning / graph mining / NLP / big data analytics/ visualization/ applications
- **Mandate**
  - Conduct basic research and development
  - Equip students with both theoretical knowledge and practical experience
- **Members**
  - Two Faculty (Prof. Aijun An, Prof. Manos Papagelis)
  - ~20 High Quality Personnel (HQP)
    - ~5 Postdoc, ~6 PhDs, ~8 MSc, ~3 Undergrads, ~1 staff

What We Are Looking For?

(solid)
Math & Stat



(solid)
Programming

(interest in)
Data Mining & ML

About you?

Contact:
Manos Papagelis
papaggel@eecs.yorku.ca
www.eecs.yorku.ca/~papaggel