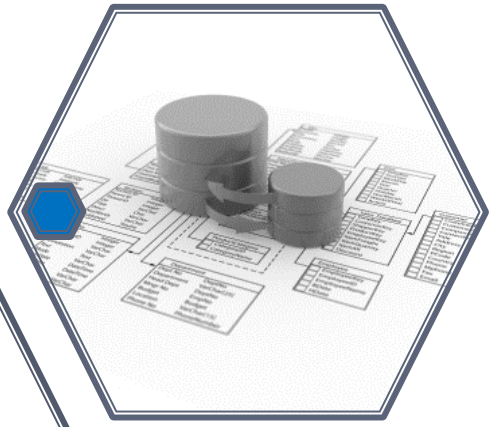
End-to-end Graph Analytics

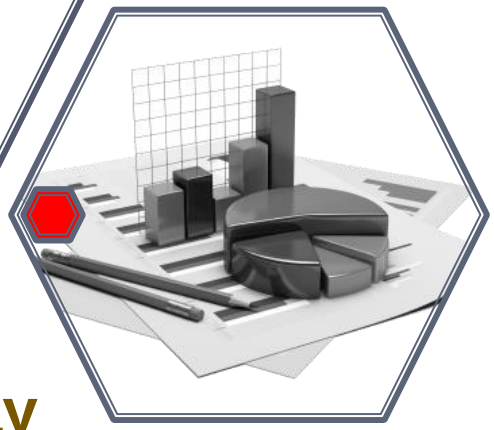
Manos Papagelis
papaggel@eecs.yorku.ca
www.eecs.yorku.ca/~papaggel

Research



Big Data & Knowledge Discovery

Data Mining

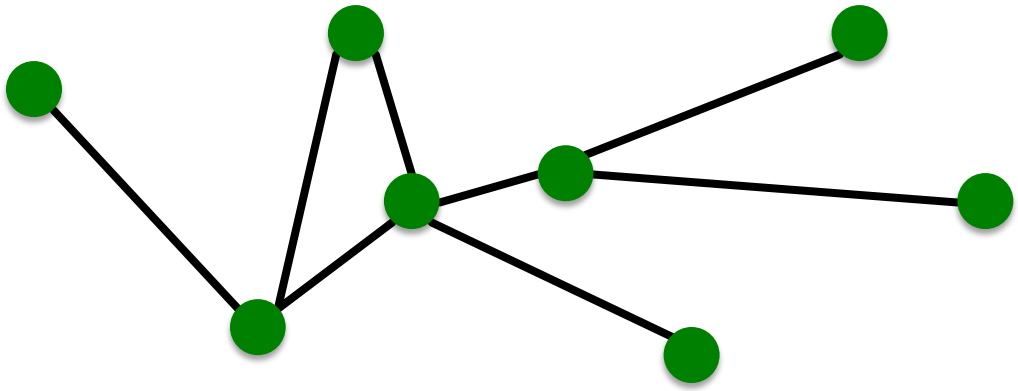


Graph Mining

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**

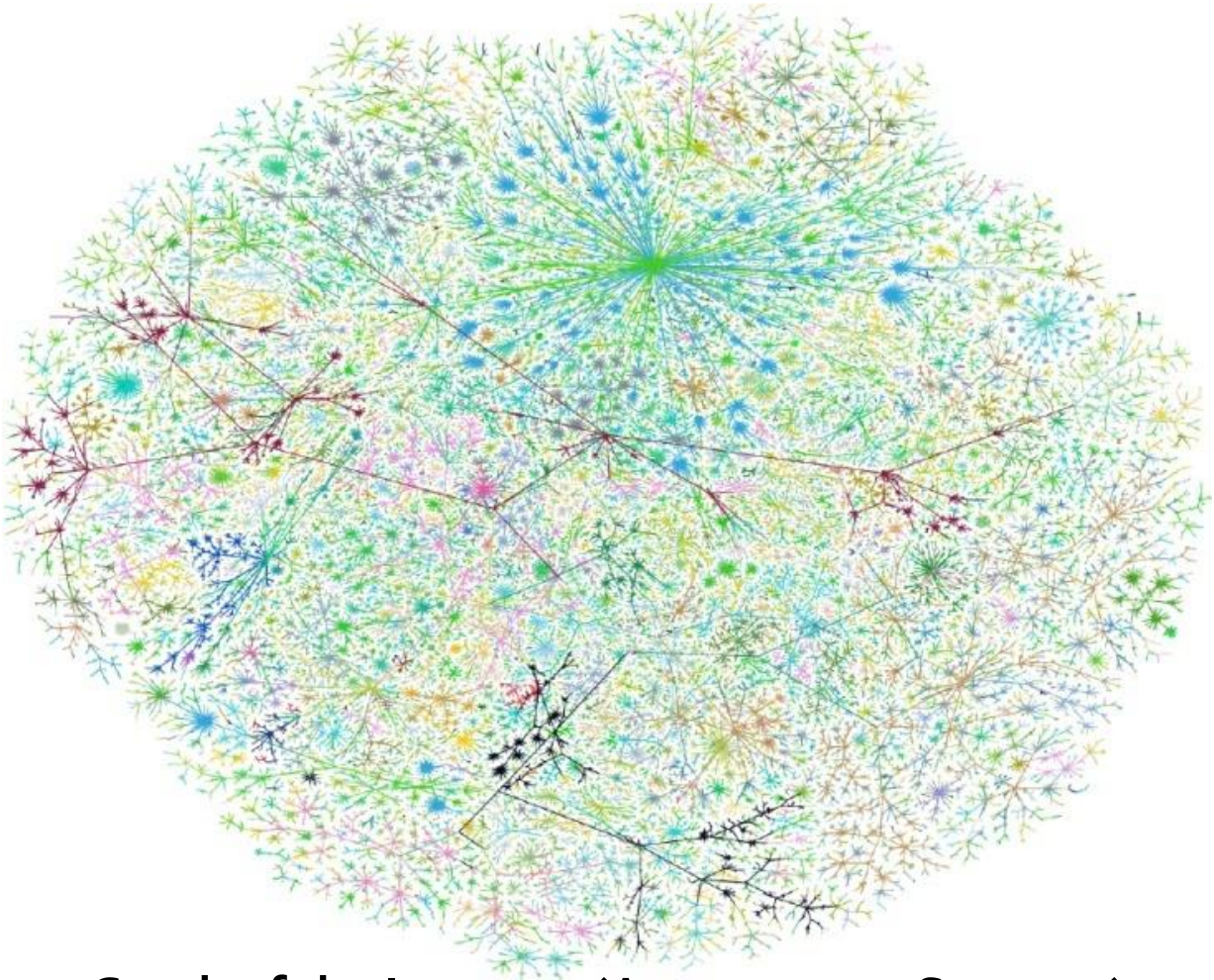
networks are ubiquitous

Networks: Social

Facebook social graph

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

Networks: Communication

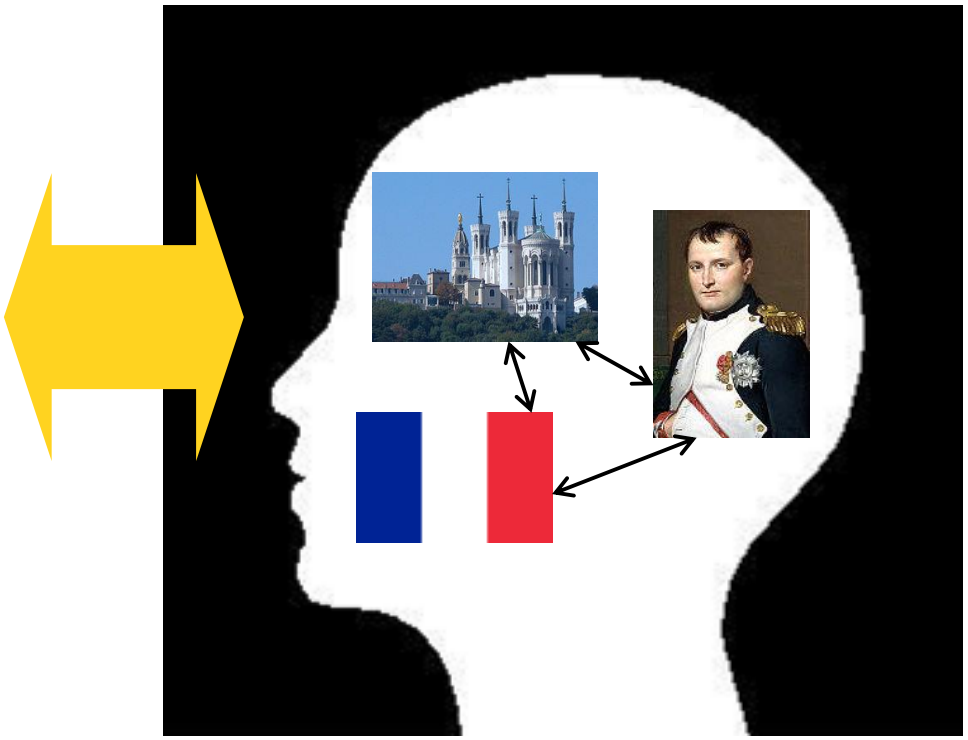


Graph of the Internet (Autonomous Systems)

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

Robustness [Doyle-Willinger, 2005]

Networks: Knowledge Graph

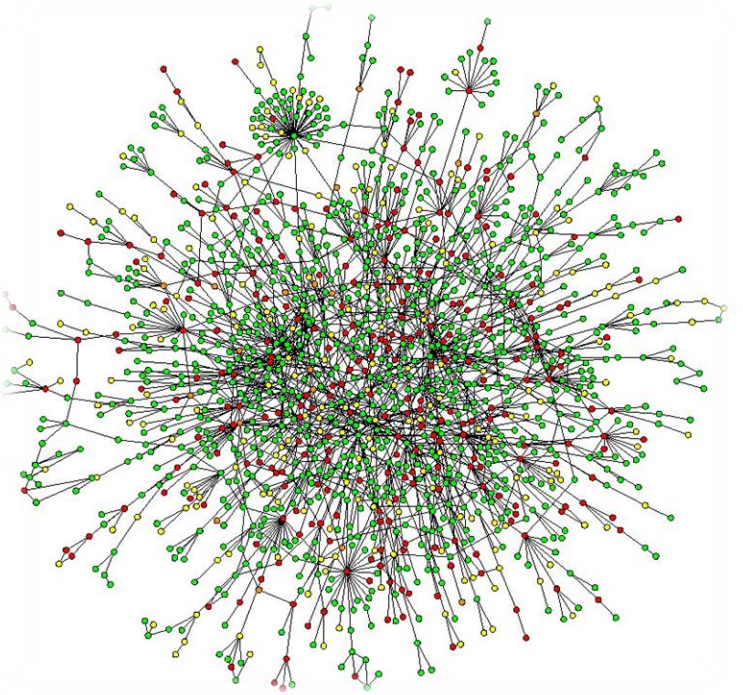


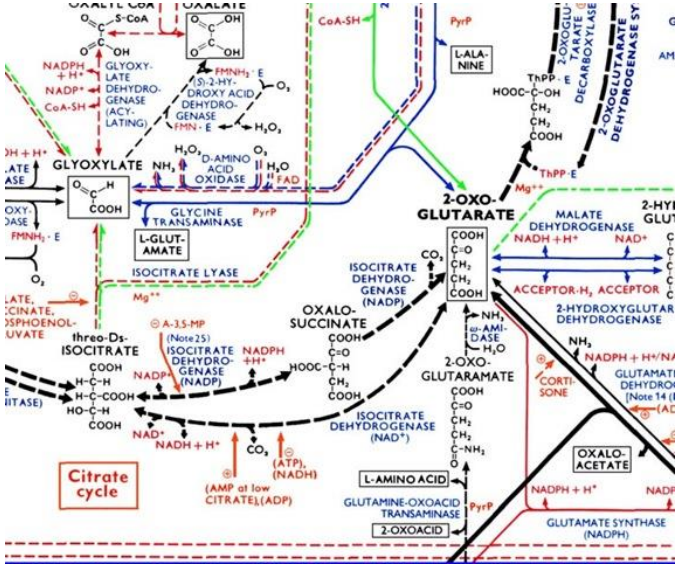
Understand how humans navigate Wikipedia

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

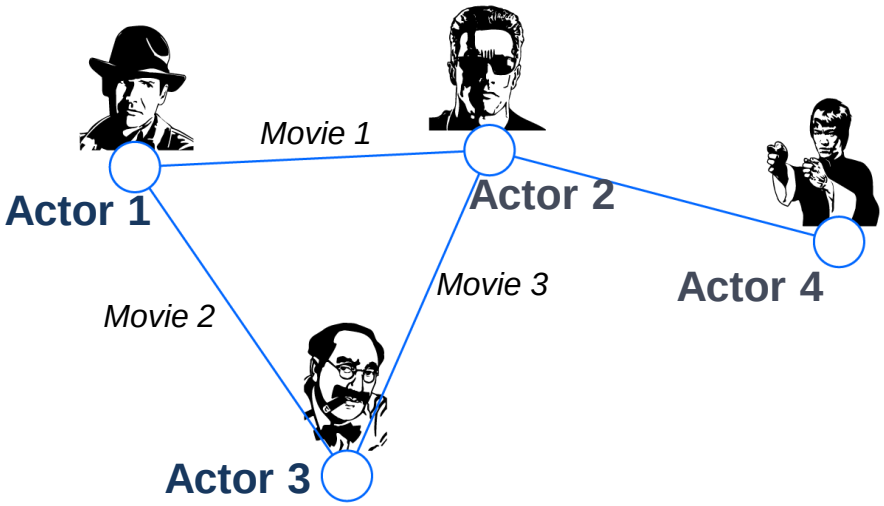
Networks: Brain

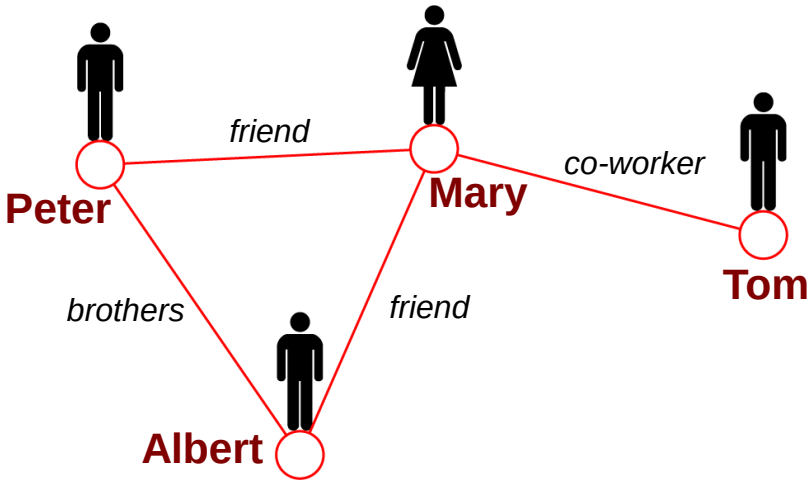
Human brain has between 10-100 billion neurons

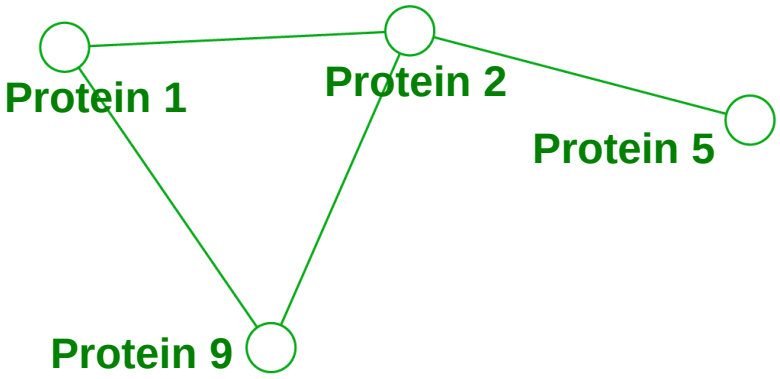
[Sporns, 2011]

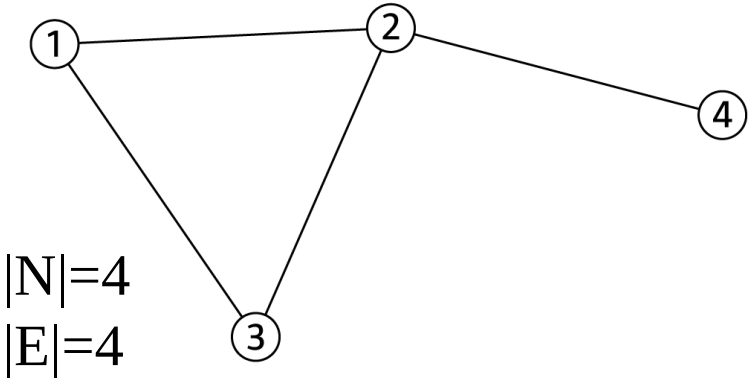
why should we care about networks?

Networks: Common Language









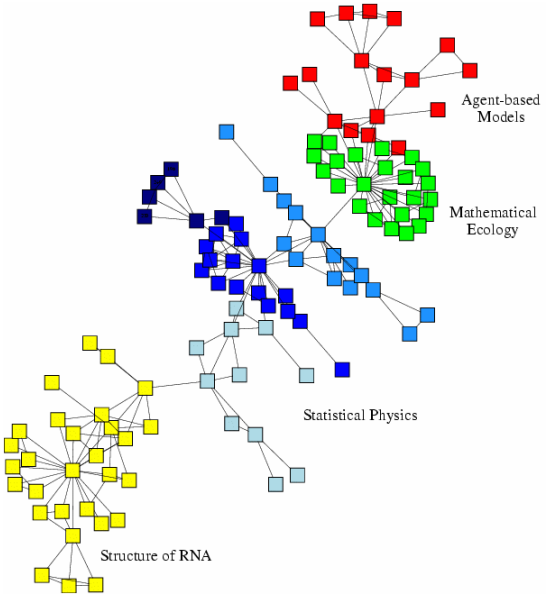
Network Analysis

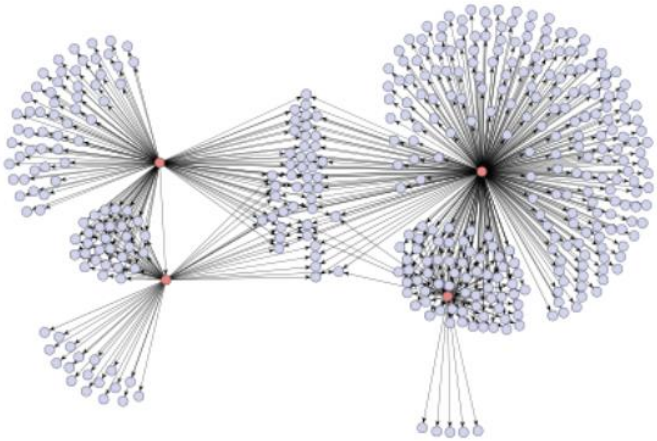
network analysis helps to reveal the *underlying dynamics* of these systems, not easily observable before

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

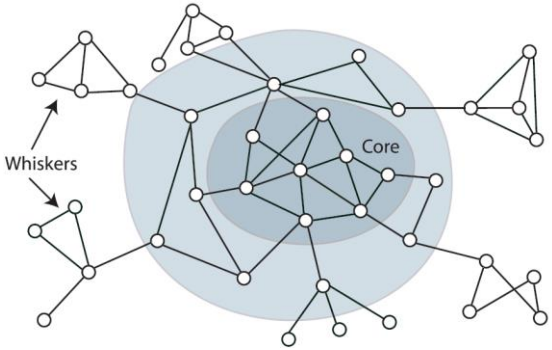
- **Algorithms**

  Methods for analyzing graphs, solving graph-related problems

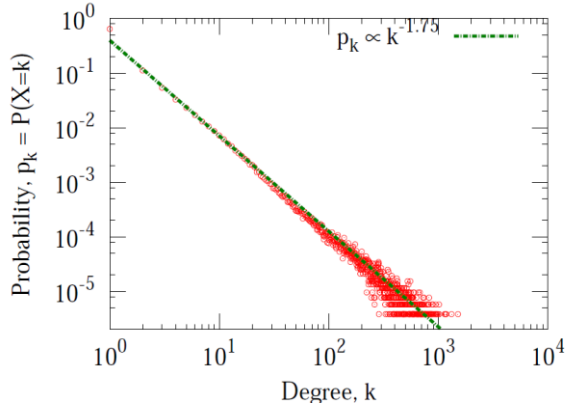
Properties

- Six degrees of separ.

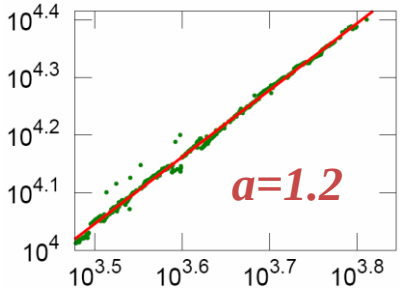
- Strength of weak ties

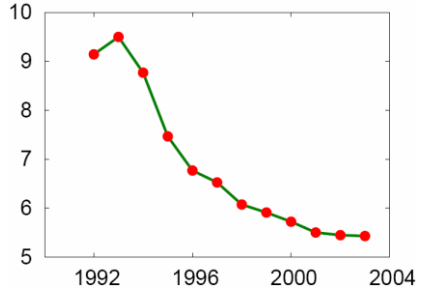


- Power-law degrees



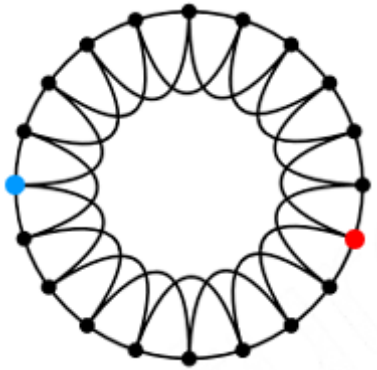
- Densif. power law, Shrinking diameter



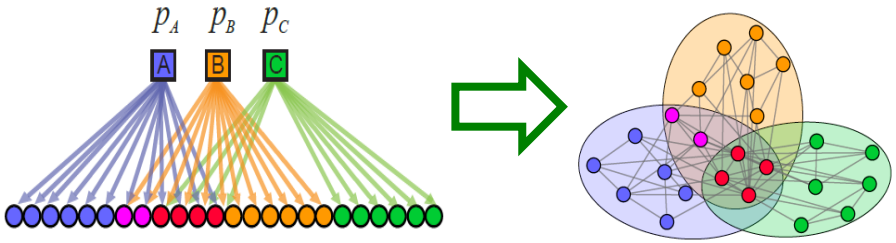


Models

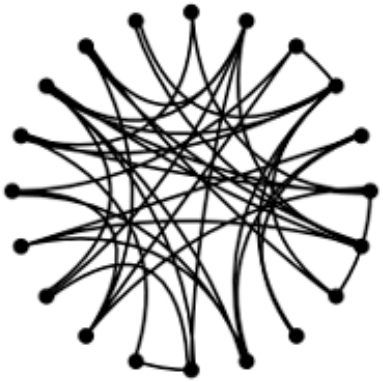
- **Erdös-Renyi model**



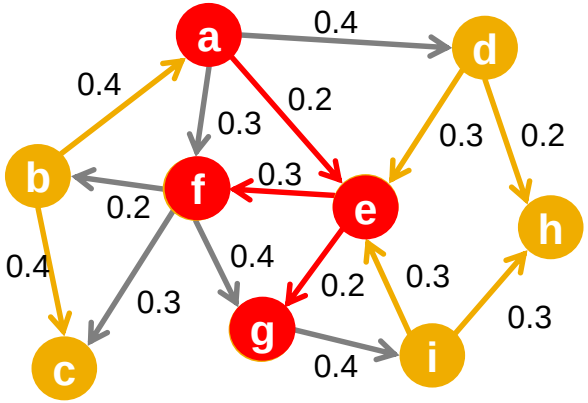
- **Community model**



- **Small-world model**

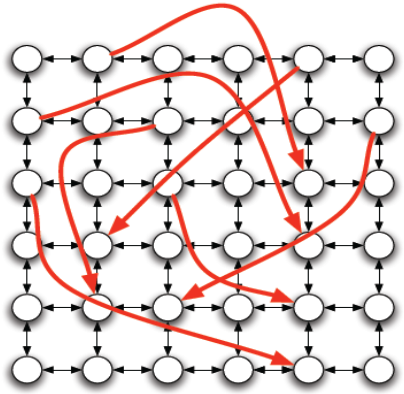


- **Cascade model**

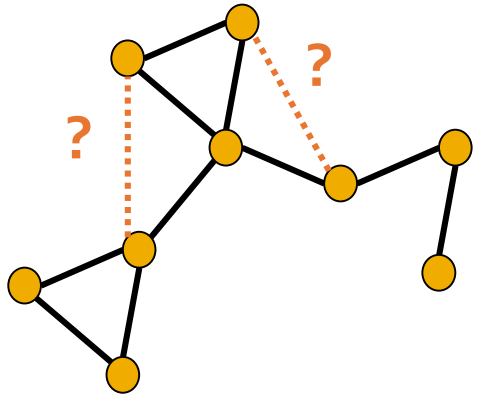


Algorithms

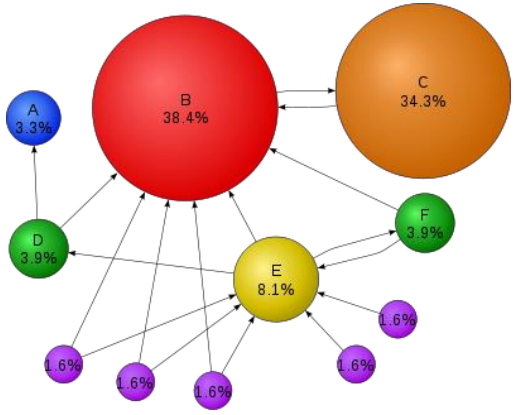
- **Decentralized search**



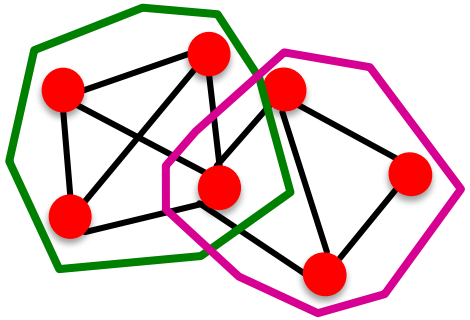
- **Link prediction**



- **Link analysis**



- **Community detection**



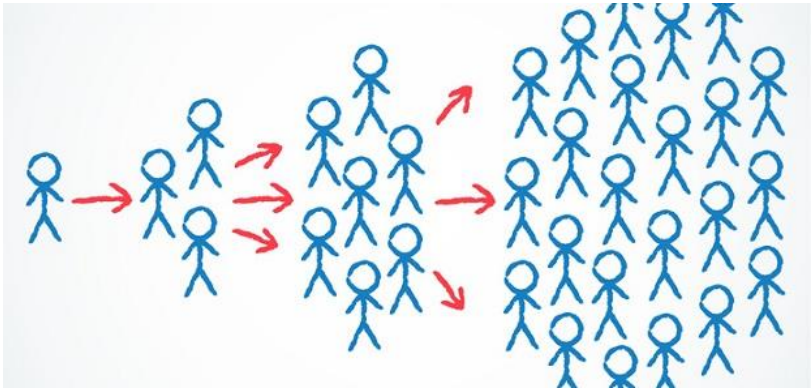
Map of Superpowers

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

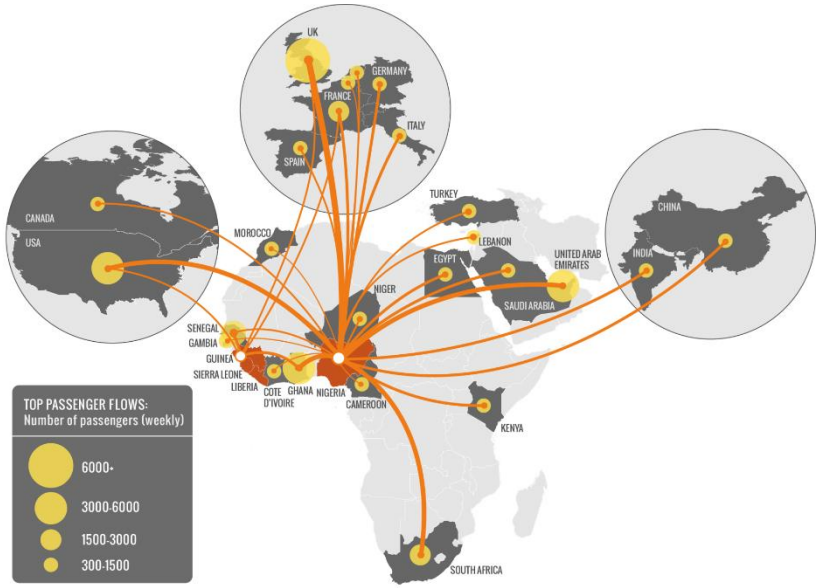
- **Social media analytics**

- **Viral marketing**

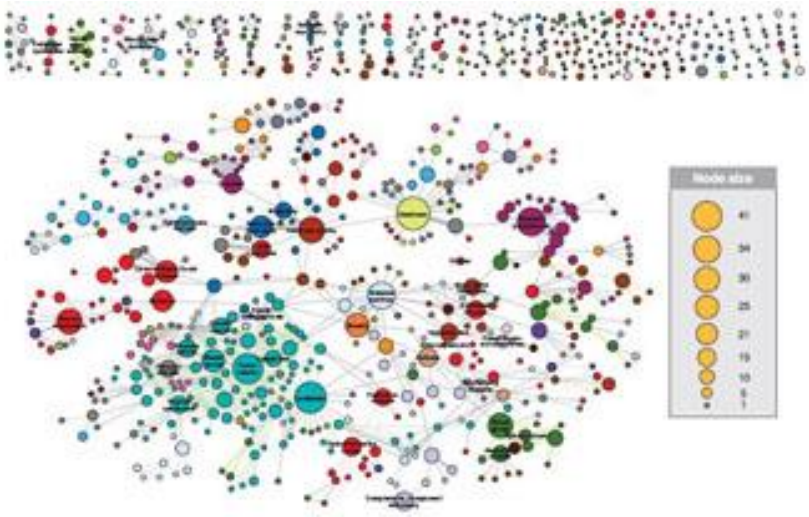


Applying Our Superpowers

- **Predicting epidemics: Ebola**



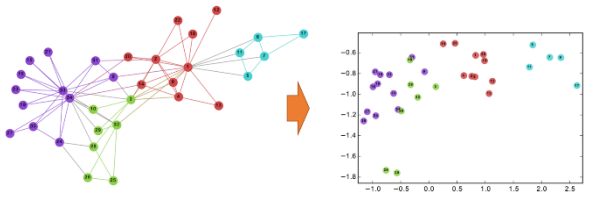
- **Drug design**



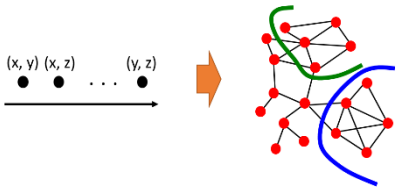
Our Research

Current Research Focus

A. Trajectory Data Mining



B. Network Representation Learning



C. Streaming & Dynamic Graphs

D. Social Media Mining & Analysis

E. City Science / Urban Informatics / IoT

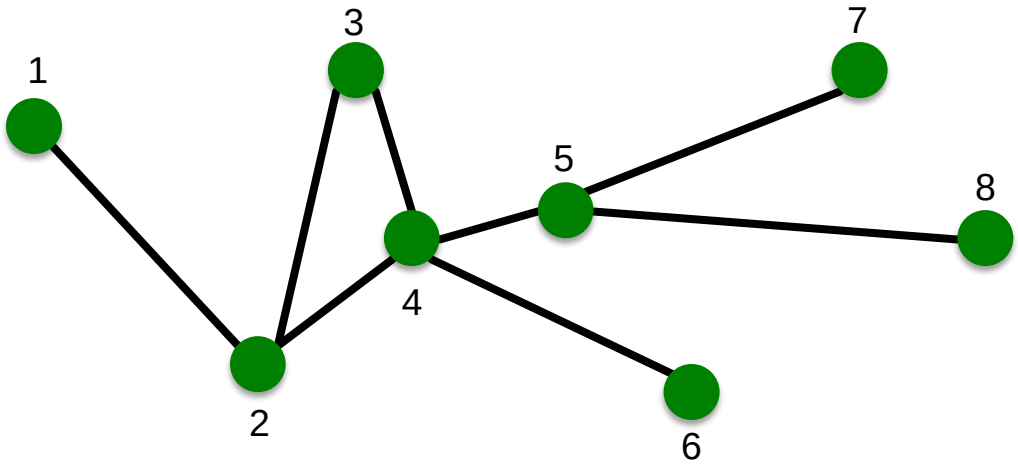
F. Natural Language Processing

Trajectory Data Mining

Problem 1: Trajectory Network Mining
Problem 2: Group Pattern Discovery of Pedestrian Trajectories

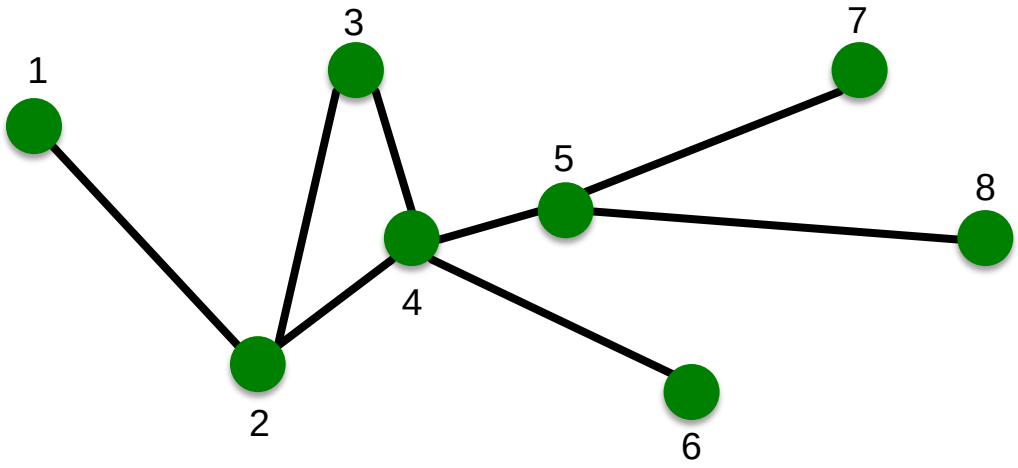
Trajectory Network Mining

Node Importance



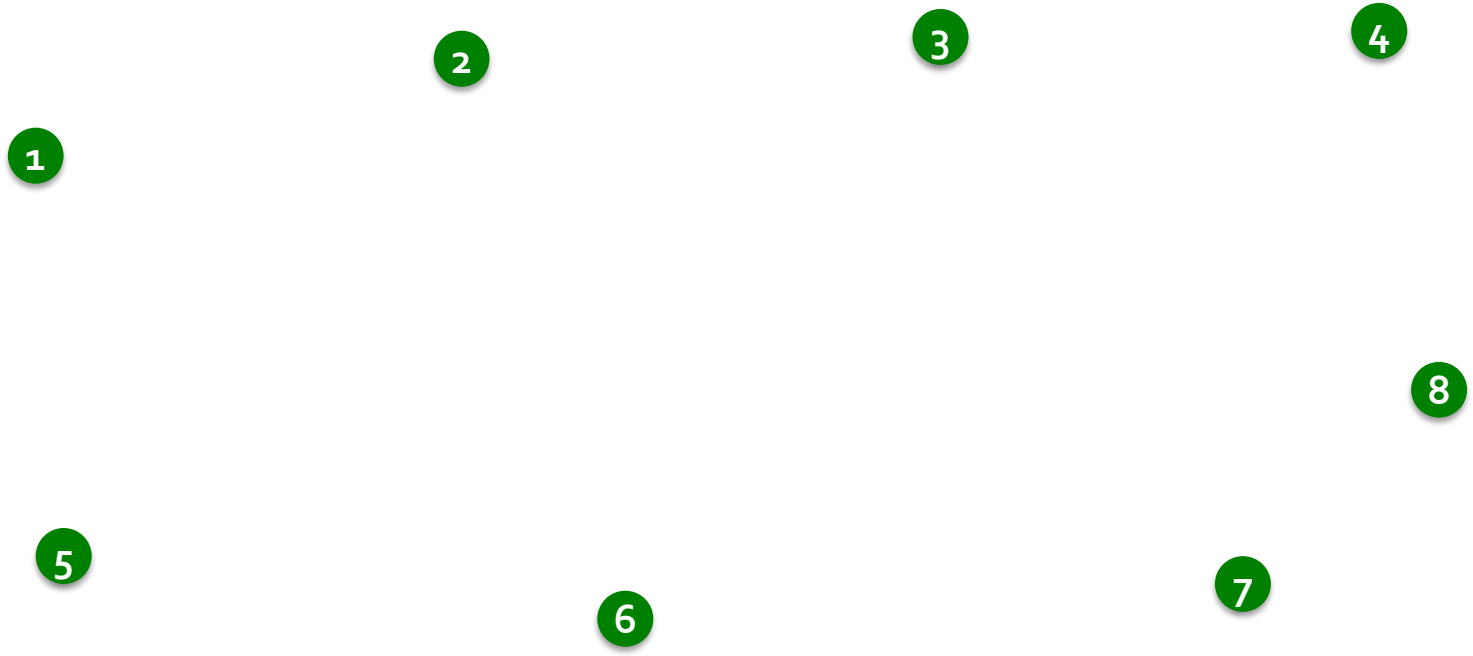
Given a **network** which node is **more important**?

Node Importance



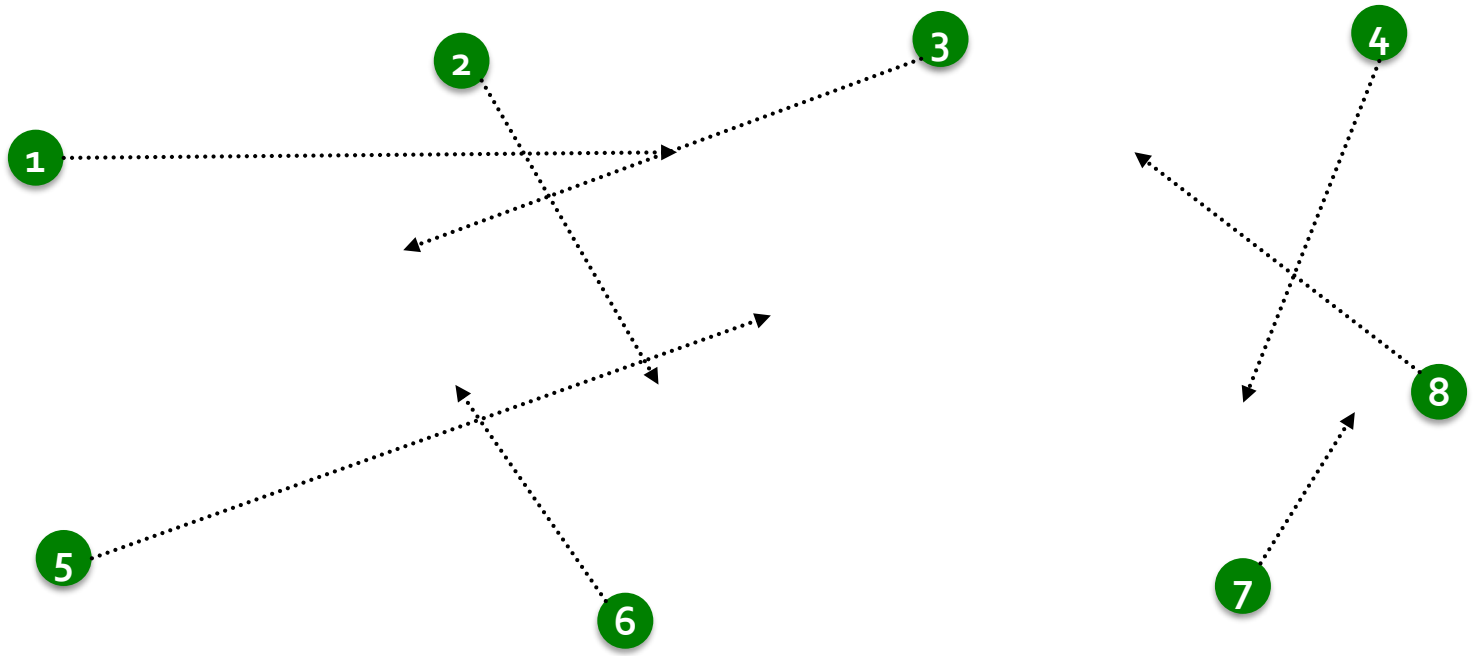
- Various notions of **node importance** (or **node centrality**)
  - **Degree centrality** (= degree of u)
  - **Betweenness centrality** (= #shortest paths passing through u)
  - **Closeness centrality** (= avg. length of shortest paths from u to all other nodes of the network)
  - **Eigenvector centrality** (= like PageRank)

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**, …

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**, …

Proximity Network at time *t*

proximity threshold θ

connect all nodes **x**, **y** where **distance(x, y) < θ**

Trajectory Networks

Trajectories

Trajectory Networks

Given a **trajectory network** which node is **more important**?

what we care about?

Node Importance in TNs

- **Node Importance in Trajectory Networks**
  - Node degrees over time (+ durations)
  - Triangles over time (+ durations)
  - Connected components over time (+ durations)
- **Applications**
  - Rich dynamic network analytics
  - Disease spreading (influenza)
  - Security (in Vehicle-to-Vehicle communications)

how to solve the problem?

Group Pattern Discovery of Pedestrian Trajectories

Pedestrian Trajectories

what is a group?

many definitions, many algorithms

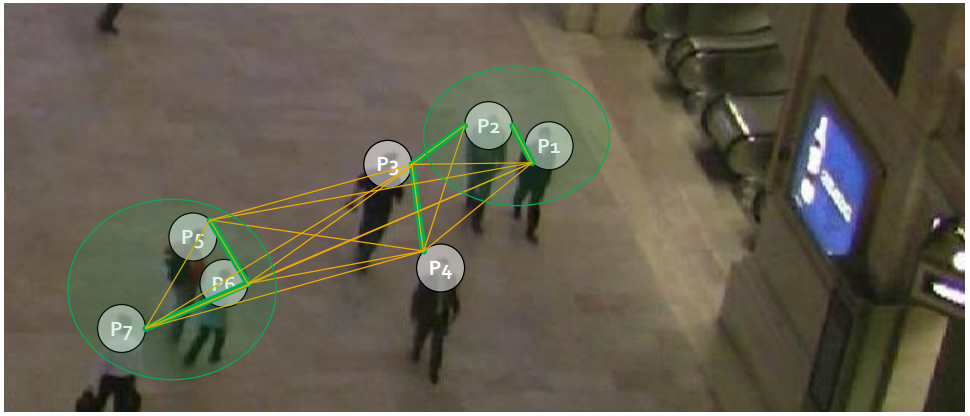
e.g., *flock*, *convoy*, *evolving-clusters*, *gathering-pattern*, … [see ACM TIST Tutorial 2015]

Finding Pedestrian Groups

Local Grouping

Intuitive method

Spatial-only



proximity threshold θ

key idea

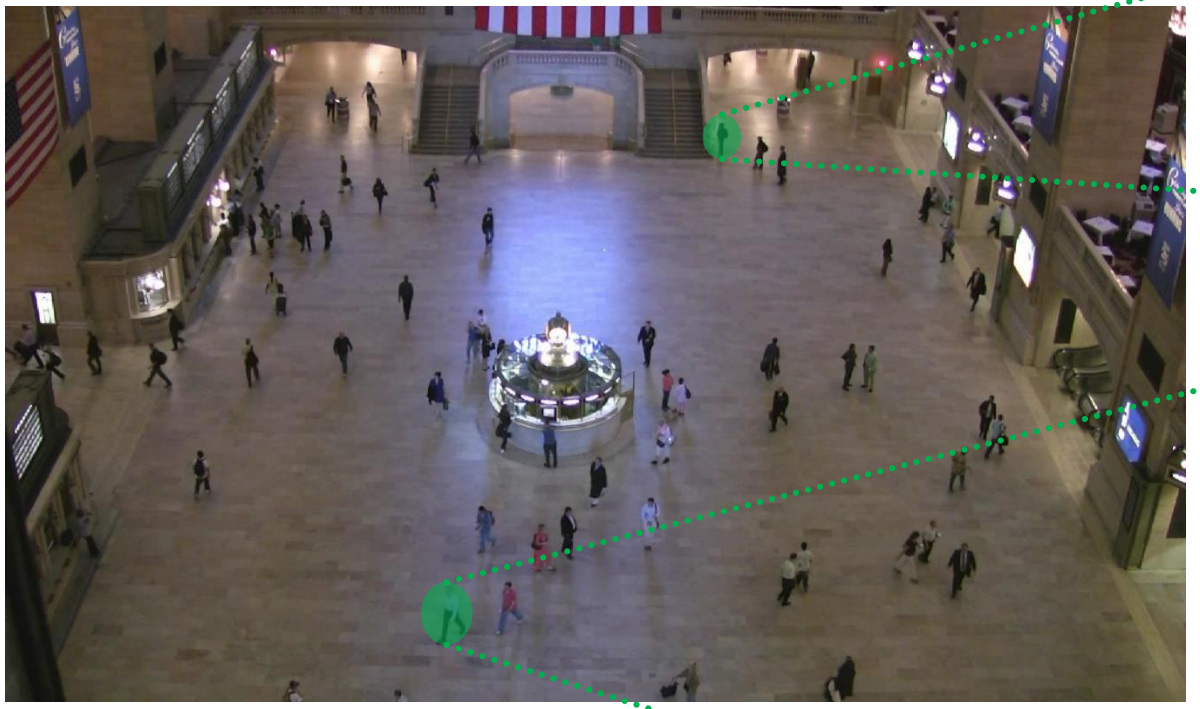
find **pairs** of pedestrians **x**, **y** where **distance(x, y) < θ**

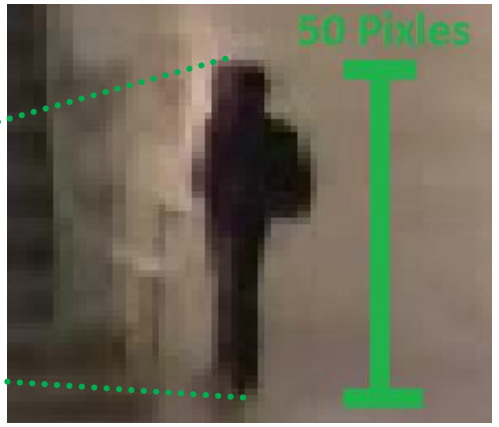
expend **pairs** to discover **groups**

Local Grouping

Challenge: Projection into Ground Plane

High perspective distortion - pedestrians closer to the camera appear larger than the ones farther away





Estimated Homography to overcome this distortion

expand the key idea to include the time dimension

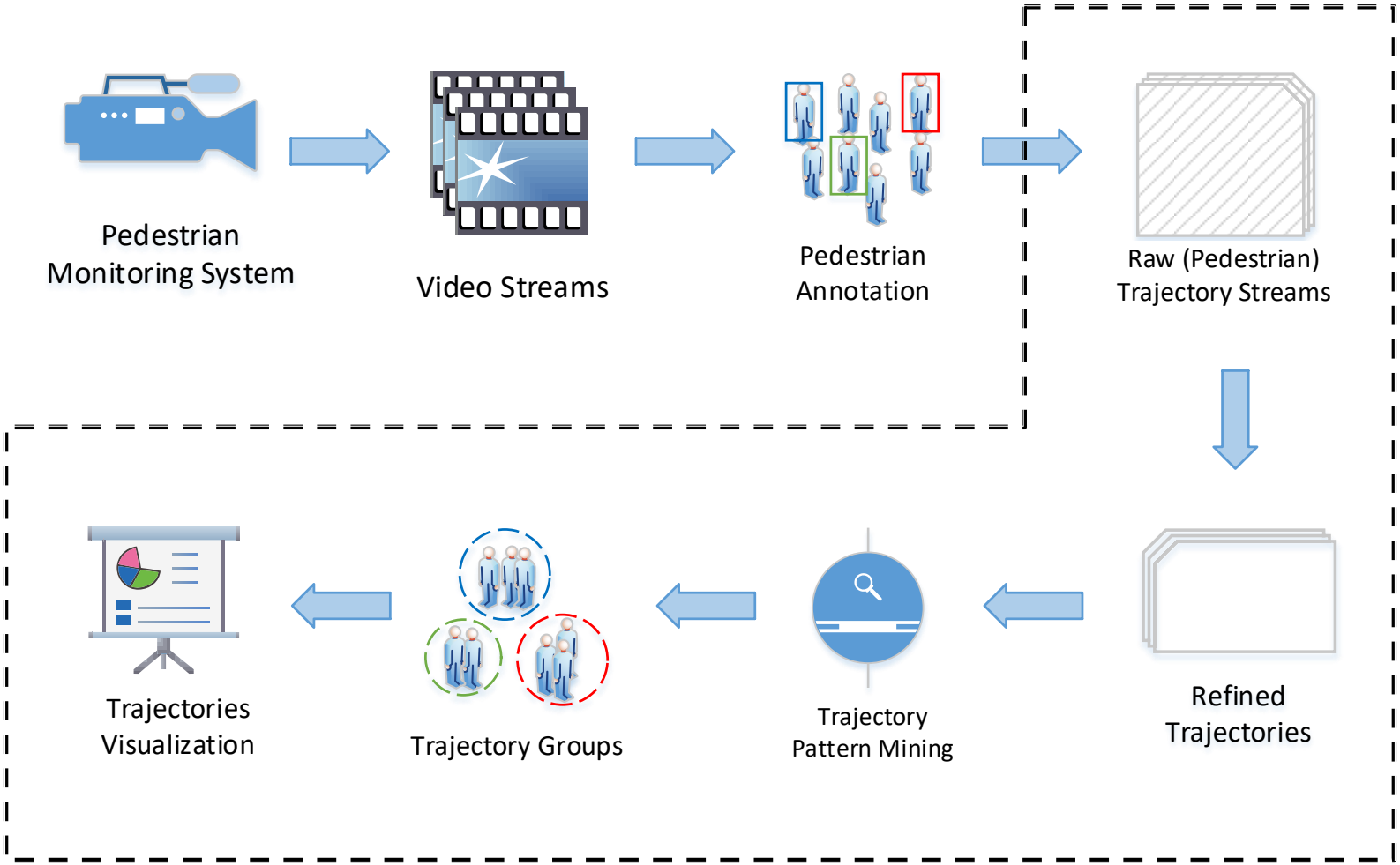
Global Grouping vs. Time-window Grouping

Global Grouping

Time-window Grouping

Trajectolizer

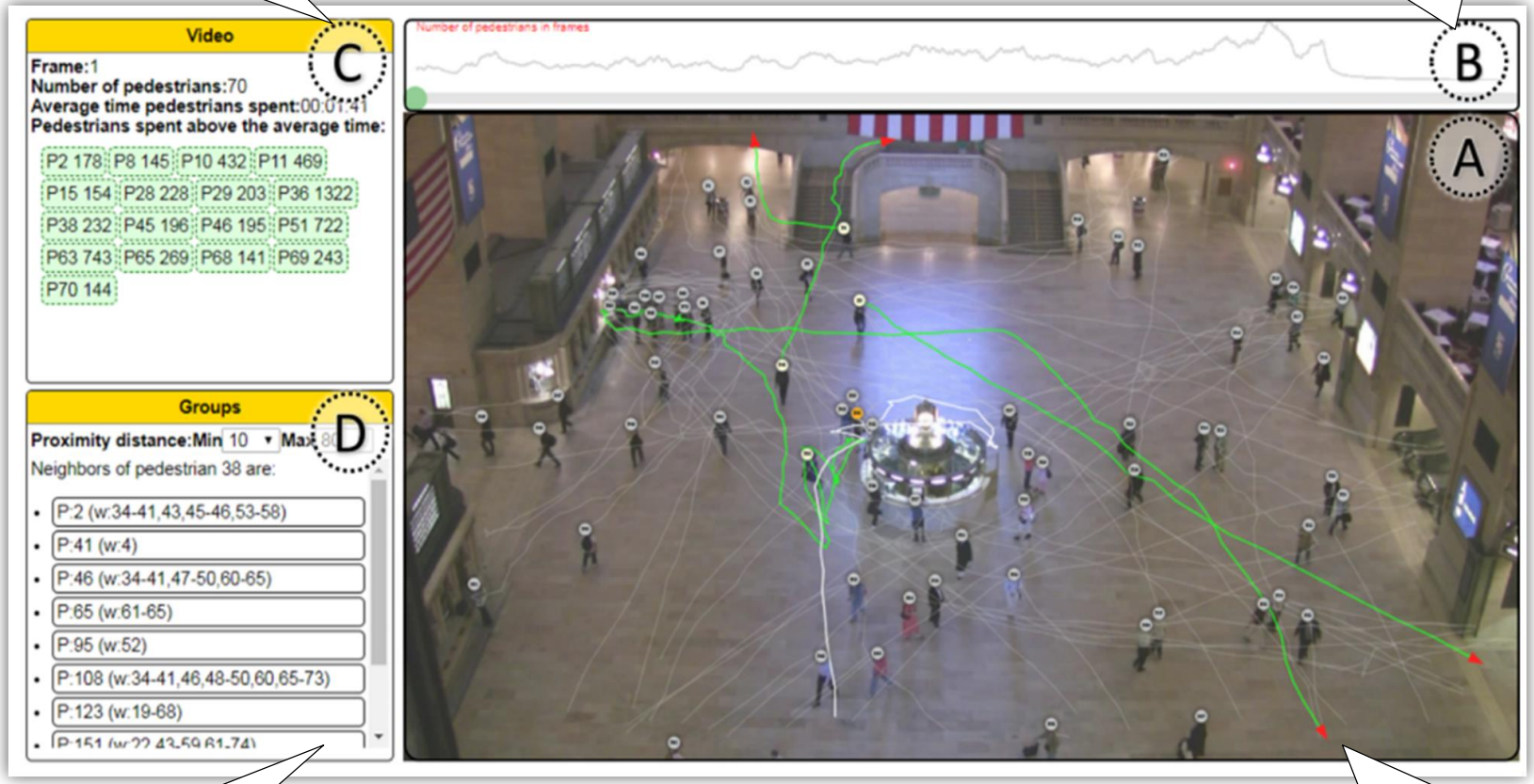
Trajectolizer: System Overview



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]

References

[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)

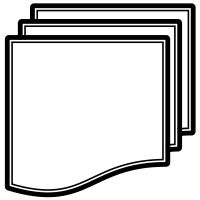
[ACM TIST Tutorial 2015] Zheng, Y. (2015). **Trajectory data mining: an overview**. ACM Transactions on Intelligent Systems and Technology (TIST), 6(3), 29.

Back to Our Long-term Goal

end-to-end graph analytics

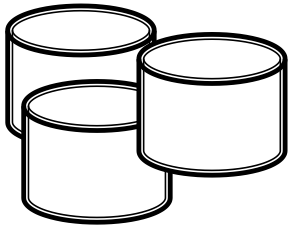
End-to-end Graph Analytics

Unstructured Data



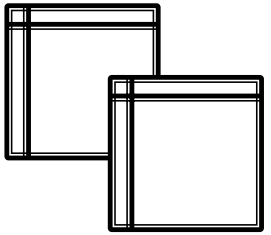
extract →

Structured Data



identify →

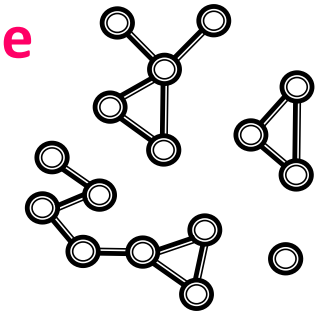
Entities/ Relationships



define →

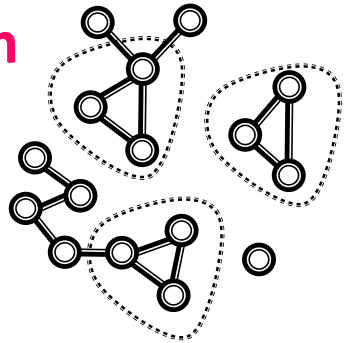
Graphs/ Networks

define →



perform →

Graph Analytics



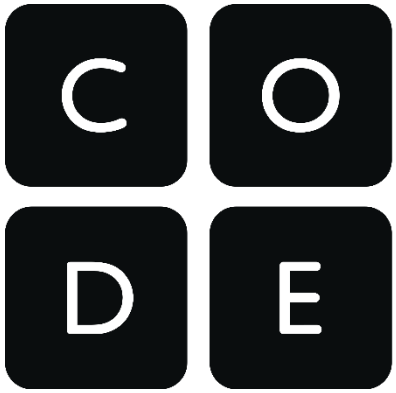
use in →

Applications/ Decision Making

Working with Us

What We are Looking For?

(solid)
Math & Stat



(solid)
Programming

(interest in)
Data Mining & ML

Thanks!

Questions?

About you?

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