Large-Scale and Multi-Structured Databases

Graph Databases Introduction

Prof. Pietro Ducange

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

UNIVERSITÀ DI PISA

CROSSLAB Innovation for industry 4.0

The Graph: Definition

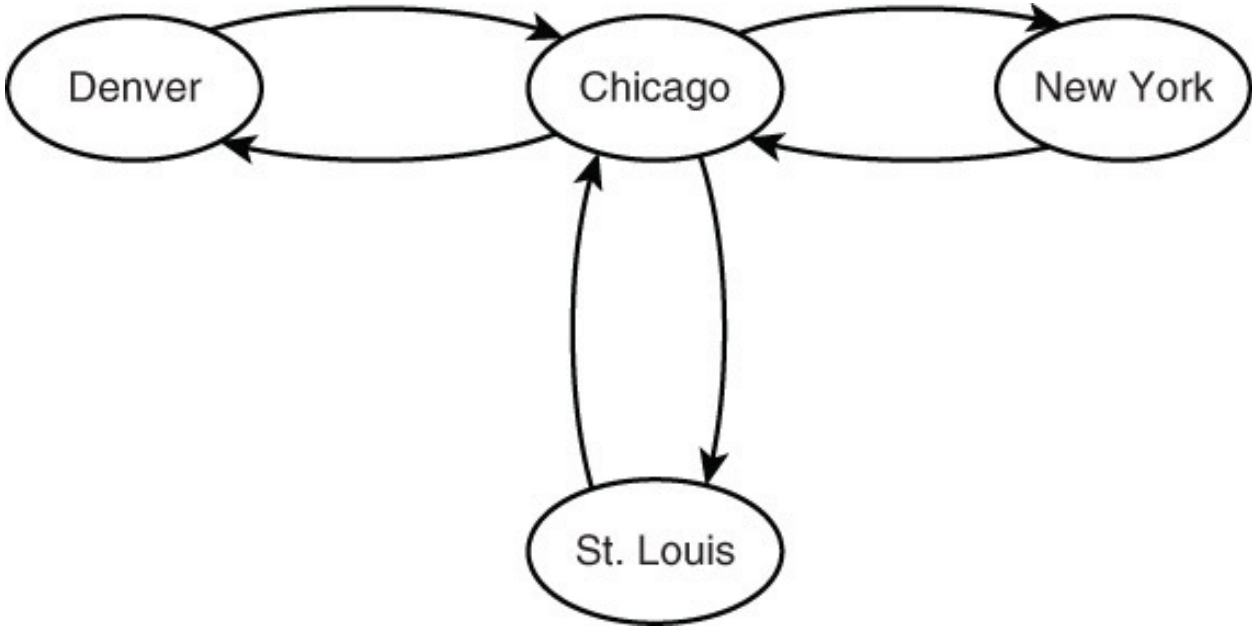
In the image below, we show an example of a simple graph, namely a collection of Vertices and Edges.

Vertex: specifies an entity. Usually, entities may belong or not to the same category. (graph nomenclature-> an entity is an instance of class/entity)

Edge: specifies a relation between two vertices. Relations may be long terms or short terms.

Vertices and edges may have ***properties***.

Graphs: Examples



Modeling Highways between Cities

Modeling Highways between Cities

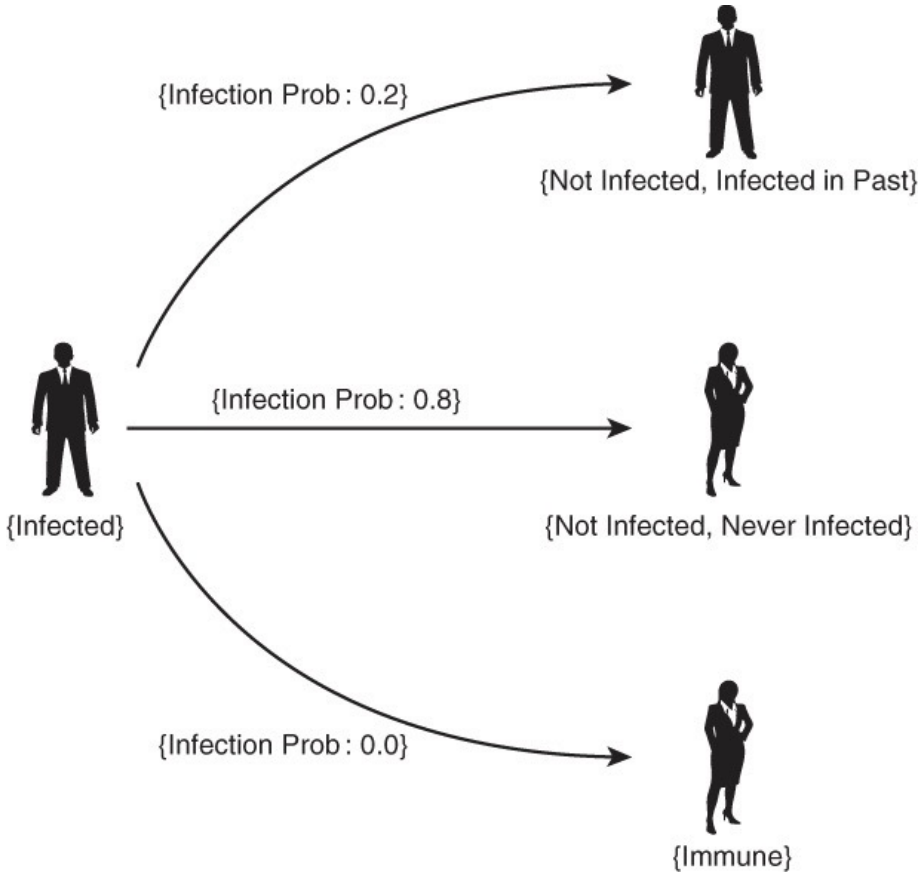
A highway could be a single edge between two cities, in which case it models the road traffic in both directions.

Alternatively, a graphical representation could use two edges, one to represent travel in each direction (for example, east to west and west to east).

Issue: Which is the "right way" to model highways?
Answer: It depends!

- ✓ If the goal is to model distance and approximate travel times between cities, then a single edge might be sufficient.
- ✓ If we are interested in more detailed descriptions of highways, such as direction, number of lanes, current construction areas, and locations of accidents, then using two edges is a better option

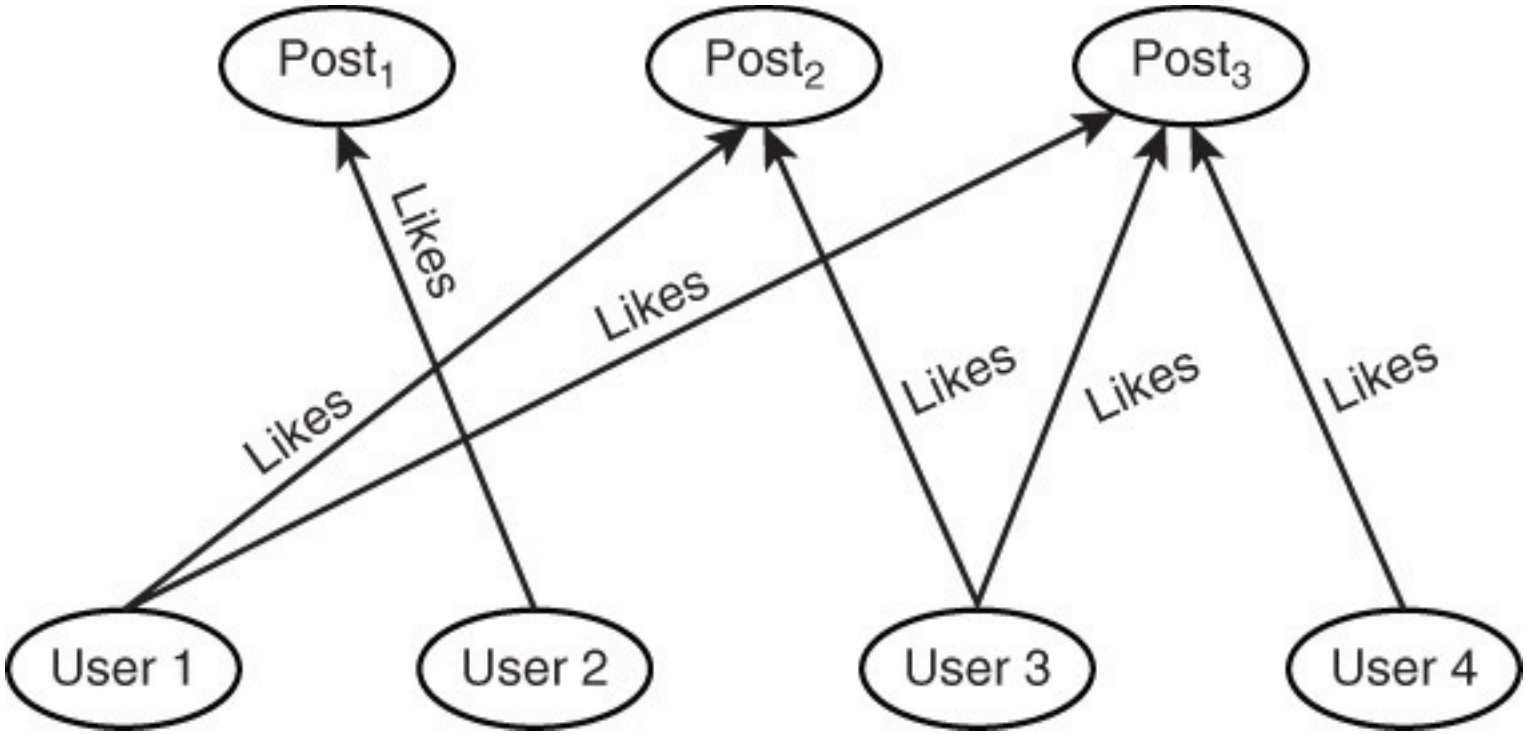
Graphs: Examples



Watch the video:
https://www.youtube.com/watch?v=lJHcg5rRAJQ

Modeling Infectious Diseases

Graphs: Examples



Modeling Social Media

Graph Databases

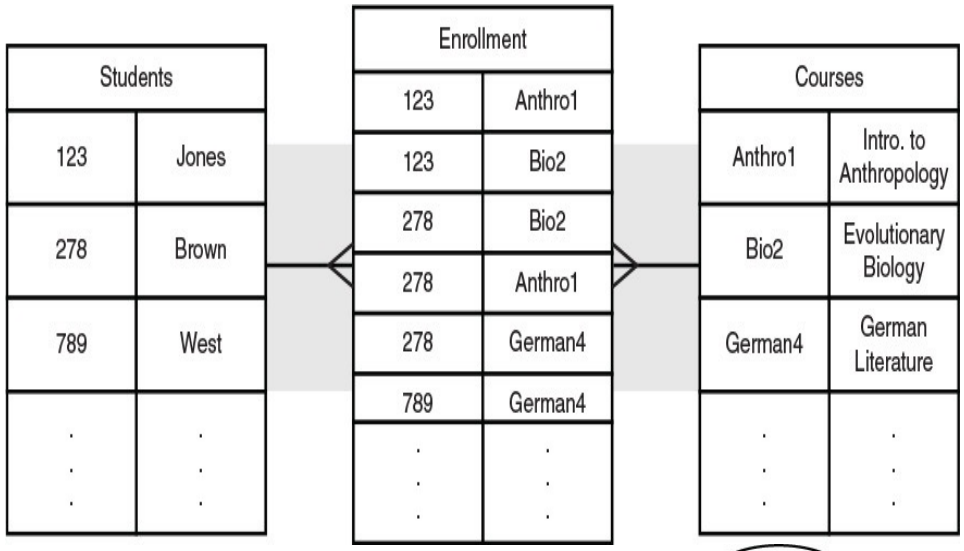
- Graph databases adopt vertices and edges to ***store*** explicit ***relations*** between ***entities***.

- In ***relational databases***, connections are not represented as links. Instead, two entities share a ***common attribute*** value, which is known as a ***key***.

- ***Join operations*** are used in relational databases to find connections or links.

- When dealing with huge amount of data (several tables with too much rows) join operation became ***computationally expensive.***

Graph Databases vs Relational Databases



Students	
123	Jones
278	Brown
789	West

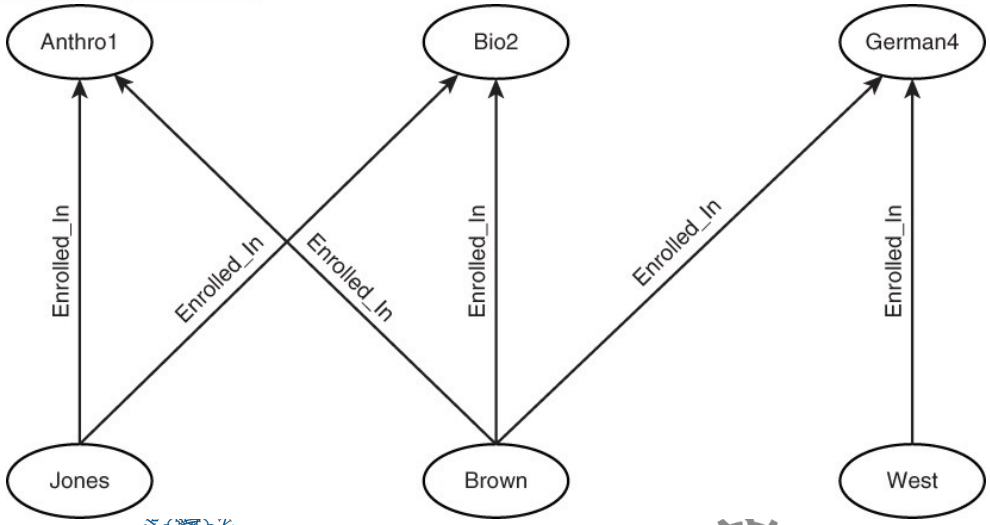
Enrollment	
123	Anthro1
123	Bio2
278	Bio2
278	Anthro1
278	German4
789	German4

Courses	
Anthro1	Intro. to Anthropology
Bio2	Evolutionary Biology
German4	German Literature

Example of query: list all the course a particular student is enrolled in.

In relational models, we need a join procedure.

The edges between students and courses allow us to ***quickly*** query the databases.



Many-to-Many Relations

Graph databases allows to modeling ***many-to-many relations*** in a easier way than relational databases.

People

U1	User1
U2	User2
U3	User3
.	.
.	.
.	.
.	.
UN	UserN

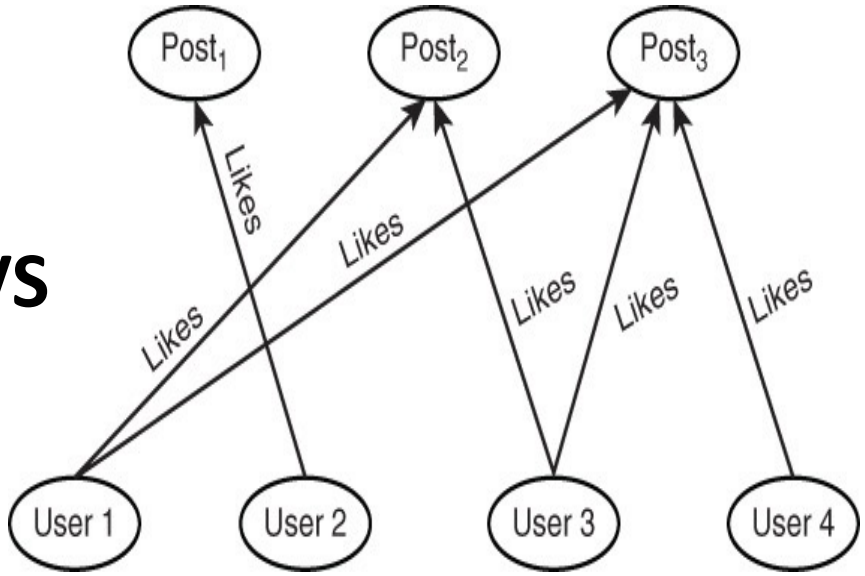
Likes

U1	P1
U1	P2
U1	P3
U2	P3
U3	PN
U3	P2
UN	P1
UN	P2

Posts

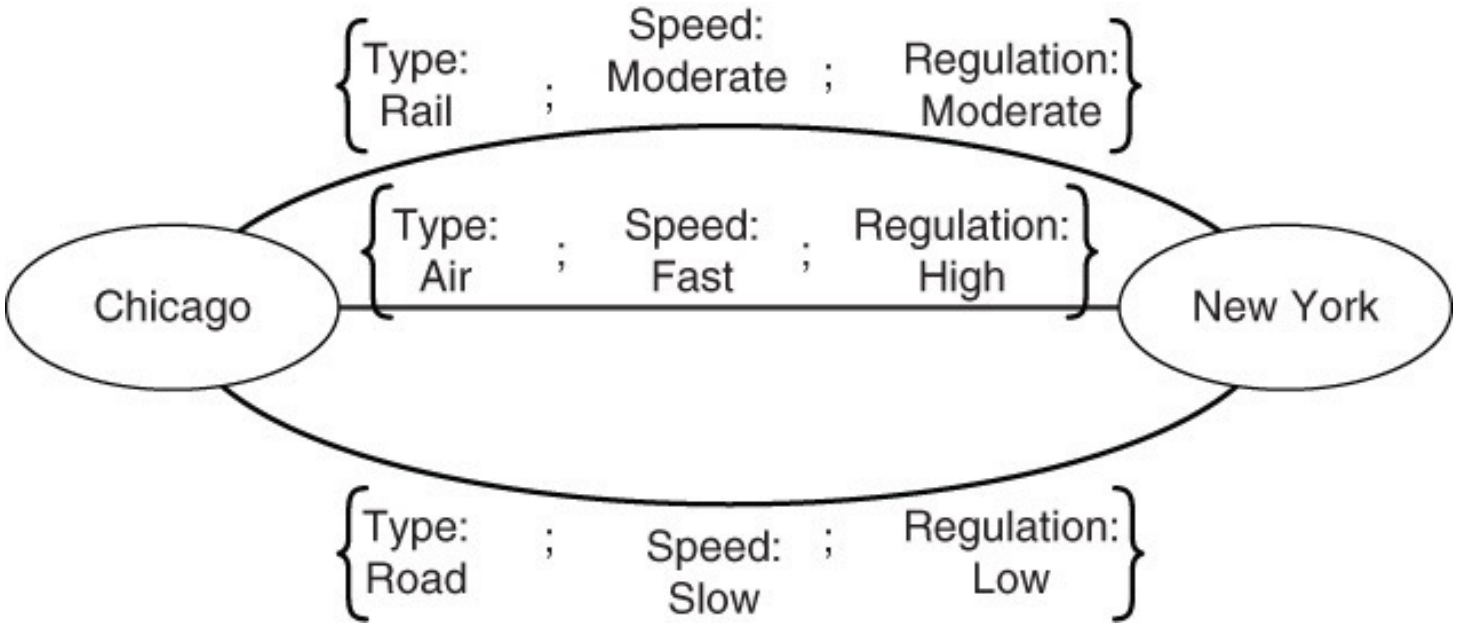
P1	Post1
P2	Post2
P3	Post3
.	.
.	.
.	.
.	.
PN	PostN

vs



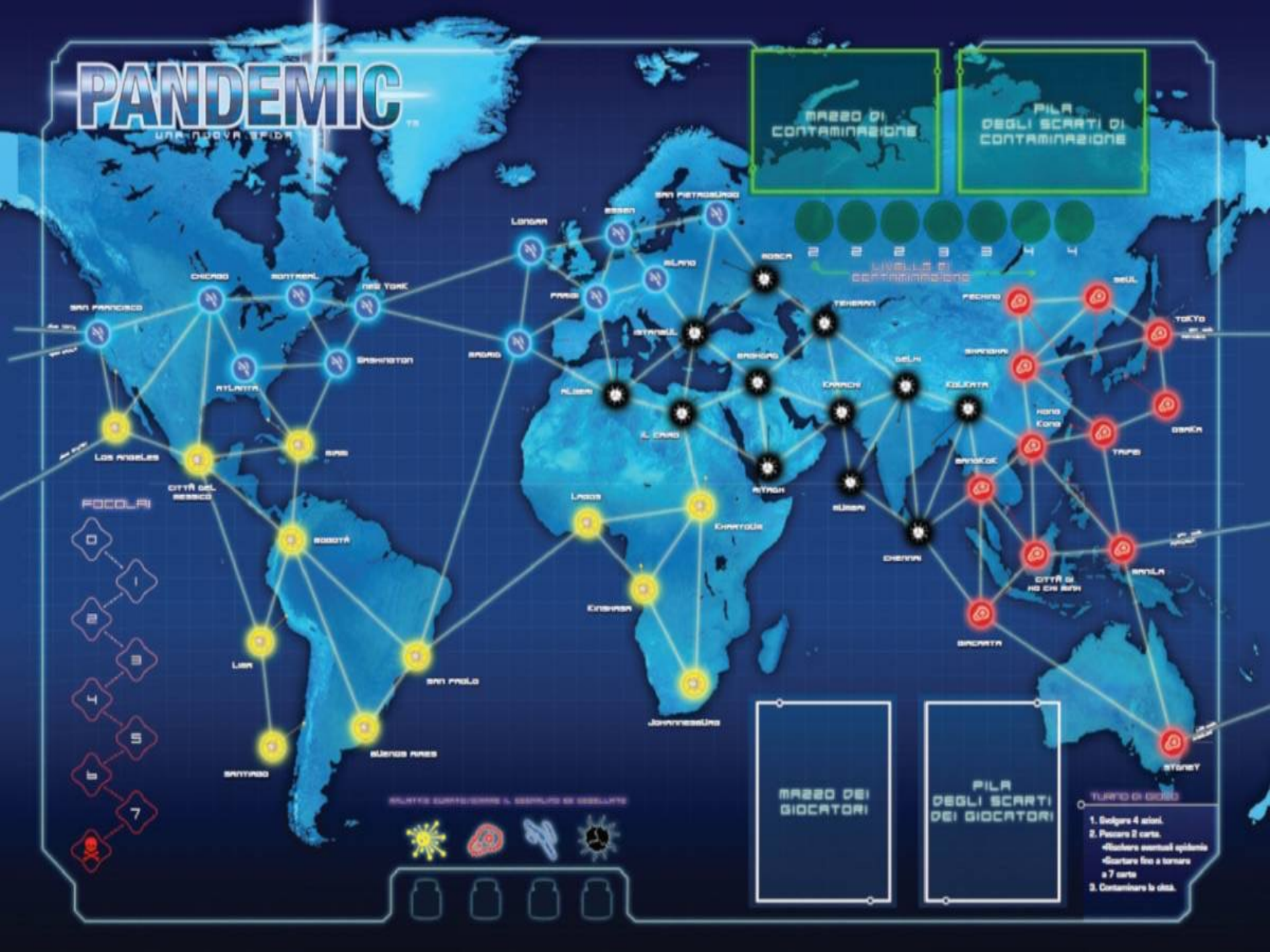
Edges allow us to ***explicitly*** modeling many-to-many relations, rather than using tables.

Modeling Multiple Relations between Entities



Different type of ***relations*** between entities may be handled exploiting ***multiple*** type of edges.

Multiple type of edges can be simply obtained specifying ***different values*** of edge ***properties***


PANDEMIC
UNA NUOVA SFIDA
MAZZO DI CONTAMINAZIONE
PILA DEGLI SCARTI DI CONTAMINAZIONE
2 2 2 3 3 4 4
LIVELLO DI CONTAMINAZIONE
FOCOLAI
0 1 2 3 4 5 6 7
MAZZO DEI GIOCATORI
PILA DEGLI SCARTI DEI GIOCATORI
TURNO DI GIOCO
1. Eseguire 4 azioni.
2. Pescare 2 carte.
• Risolvere eventuali epidemie
• Scartare fino a tornare a 7 carte
3. Contaminare le città.

Pandemic Game

Graph-Based Representation:

- Vertices (Nodes): Cities represented as nodes.
- Edges (Lines): Connections between cities indicate pathways for disease transmission.

Node Properties:

- Number of connections (edges) with other cities.
- Infection level indicated by colored cubes.

Edge Properties:

- Connects two infected cities, two healthy cities, or a mix of one infected and one healthy city.

Pandemic Game

Objective of the Game:

- Collaborate to contain and eradicate four infectious diseases worldwide.
- Players win by eradicating all diseases or losing if outbreaks overwhelm the board.

Game Components:

- Board: Displays major global cities, each connected by paths (edges).
- Disease Cubes: Represent infection levels in cities.
- Player Roles: Each player has unique abilities to aid in eradicating infections.

Suggested Readings

Chapter 12 of the book "*Dan Sullivan, NoSQL For Mere Mortals, Addison-Wesley, 2015*".

Watch the Pandemic Game video:

https://www.youtube.com/watch?v=63Ha1ktxvoY&ab_channel=Geek%26Sundry

Images

If not specified, the images shown in this lecture have been extracted from:

"Dan Sullivan, NoSQL For Mere Mortals, Addison-Wesley, 2015"