Mining Graph Data

Karel Vaculík and Luboš Popelínský

Knowledge Discovery Lab
Faculty of Informatics, Masaryk University Brno
Czech Republic

popel@fi.muni.cz www.fi.muni.cz/~popel

Is there a need for mining in graphs?

or exist already tools

that can manage it?

Movie information as a graph

IMDb

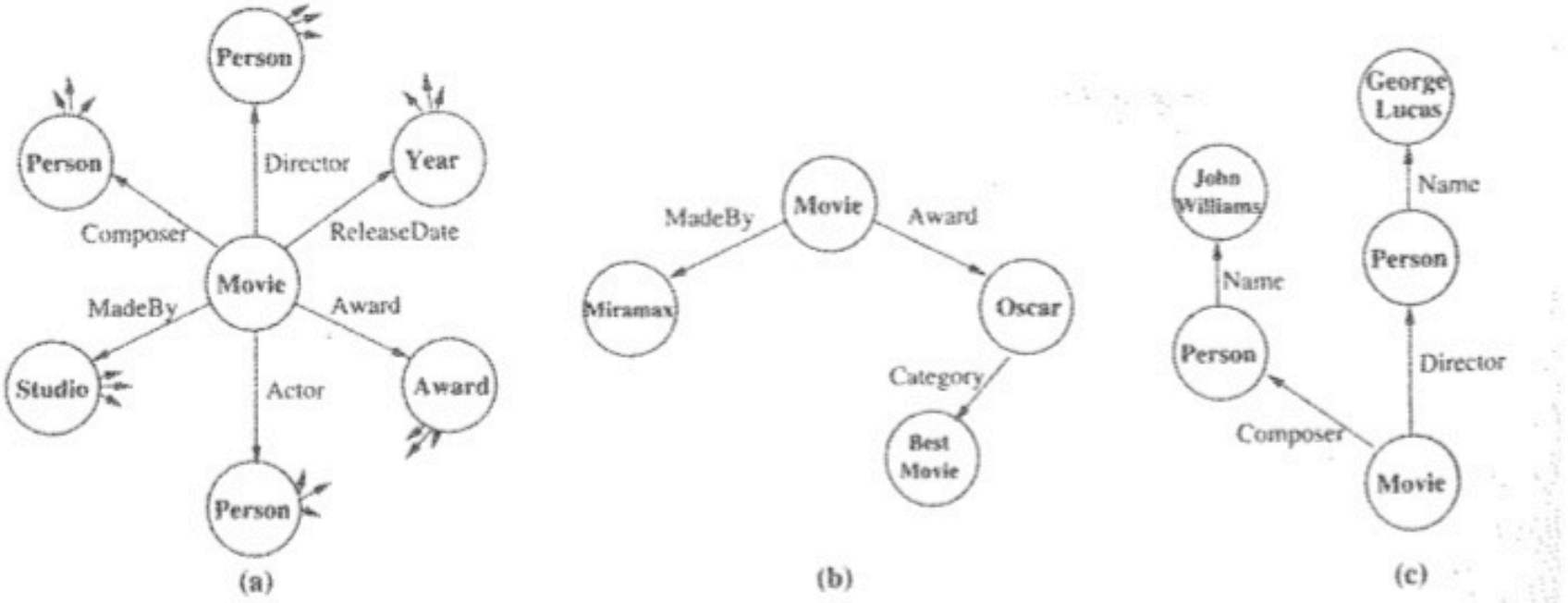


Figure 1.1. (a) Possible graph representation for information related to a single movie. (b) One possible frequent subgraph. (c) Another possible frequent subgraph.

Movie information as a graph

What commonalities can we find about movies in IMDb?

by frequent subgraph discovery:

Movies receiving awards (Oscars, Golden Globes) come from the same small set of studios

Certain director/composer pairs work frequently together

Movie information as a graph

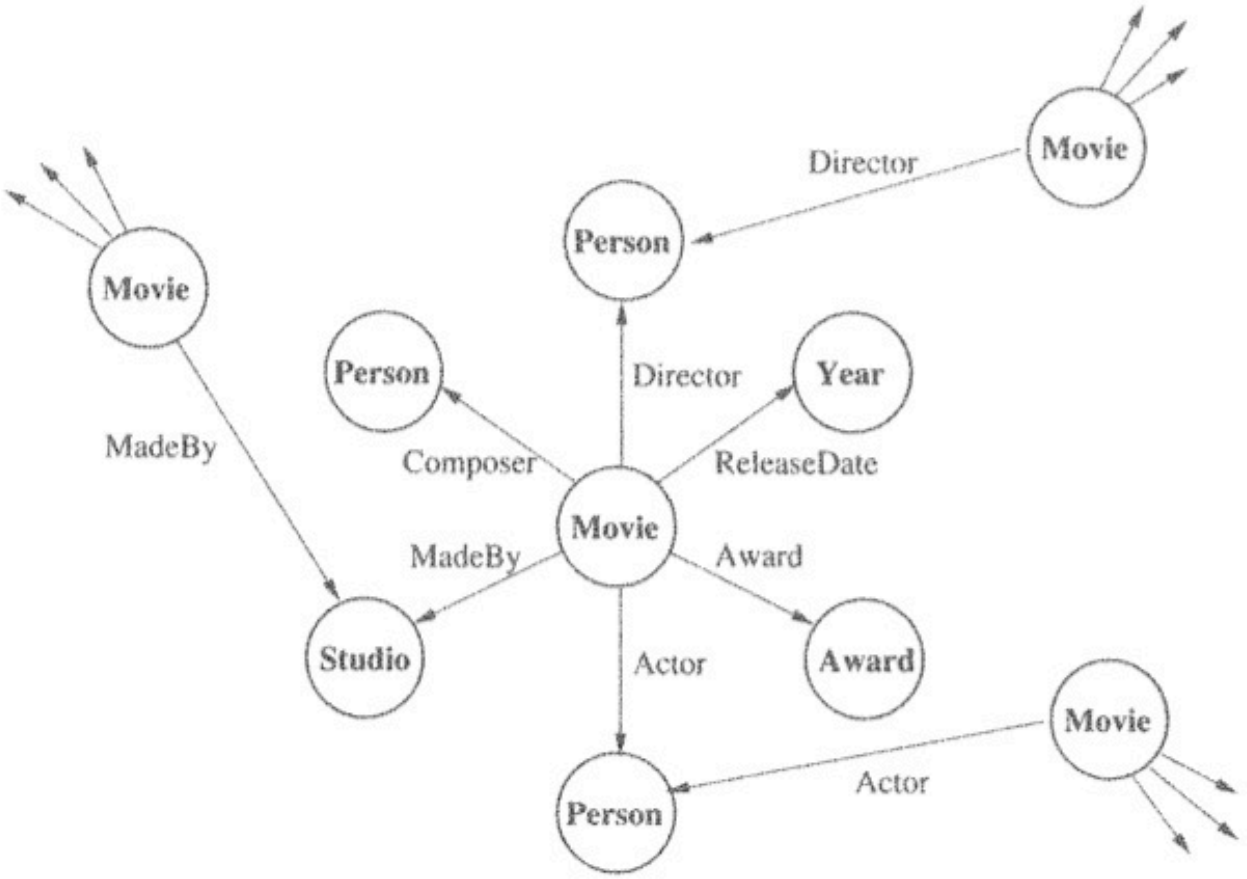


Figure 1.2. Second graph representation in which relationships between data points are represented using labeled edges.

Movie information as a graph

What common relationship can we find between object in the db?

Movies made by the same studio also have the same producer.

An emerging film star may be characterized by a sequence of successful movies.

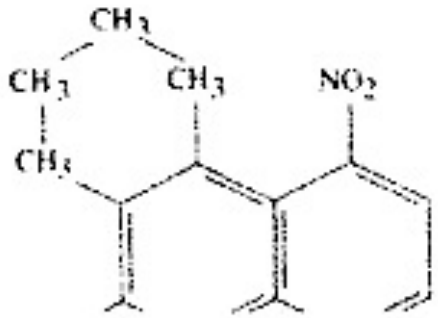
Will a movie make more than $2 million in its opening weekend?

Will be the movie nominated for an award? but also

for inferring missing links in a movie graph

Mutagenesis data

mutagenic vs. non-mutagenic substances

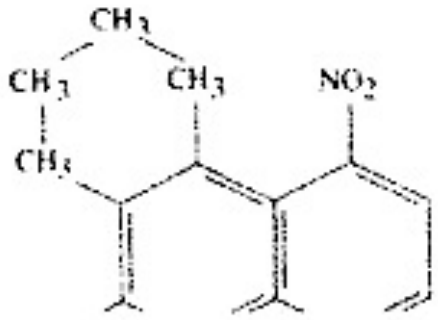


What are commonalities for each of those classes

What are commonalities for each of those classes,
e.g. subgraphs, that distinguish
mutagenic and non-mutagenic substances?

Mutagenesis data

Inductive logic programming can help.



atom(Id, Element, AdditionalInfo, ….).

bond(Id1, Id2, Arity, AdditionalInfo, …).

ring(…).

Mutagenesis data

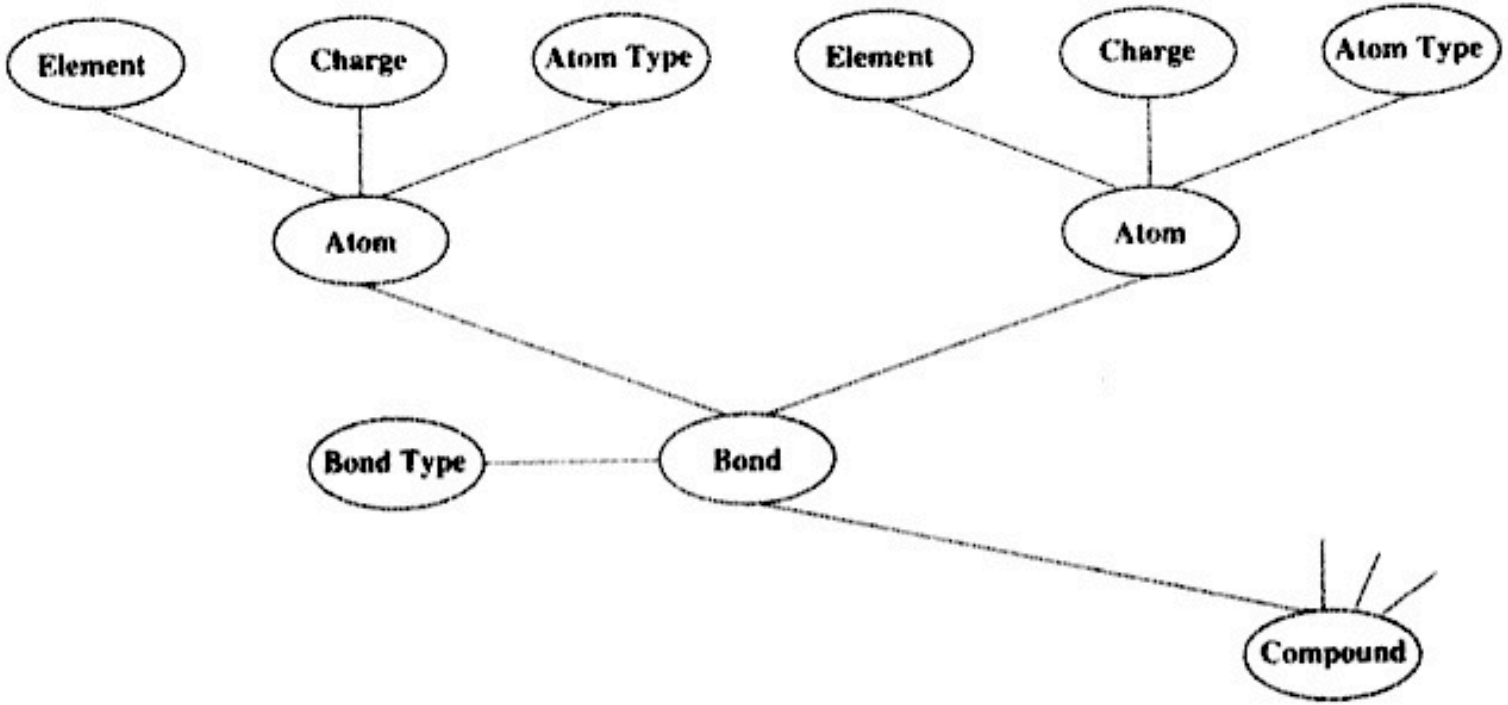


Figure 1.4. Graph representation for a chemical compound.

Web

Web mining =

web usage mining

web structure mining

web content mining