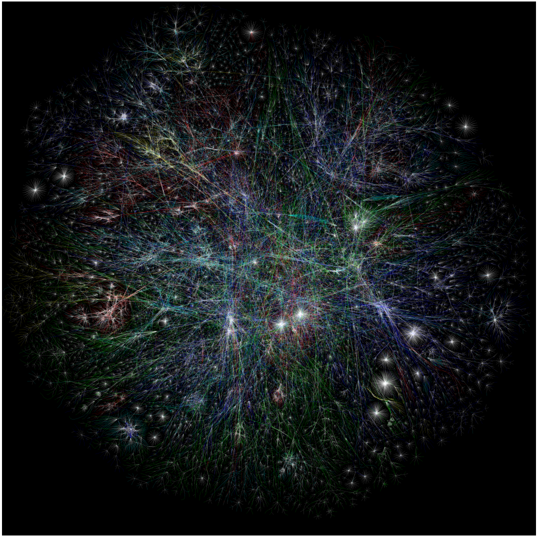
Learning and Reasoning with Graph Data: Neural and Statistical-Relational Approaches

Graph Data: Representation and Reasoning

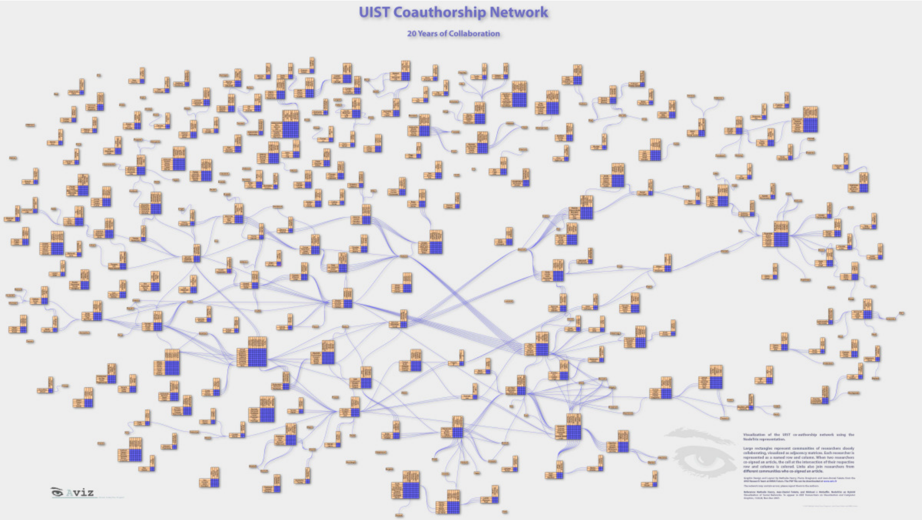
Manfred Jaeger

Aalborg University

Networks!

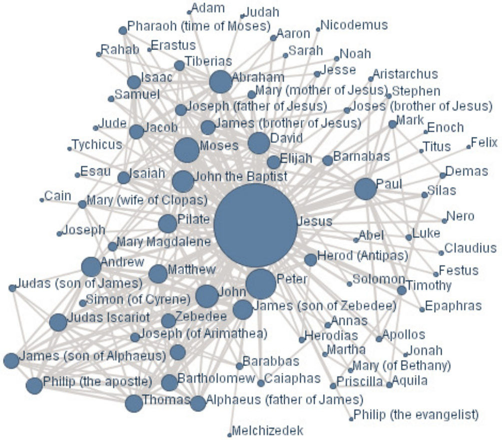


[https://www.opte.org/the-internet]



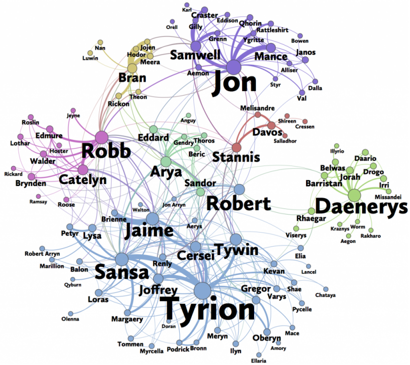
[www.aviz.fr/old/gallery/]

The New Testament social network:

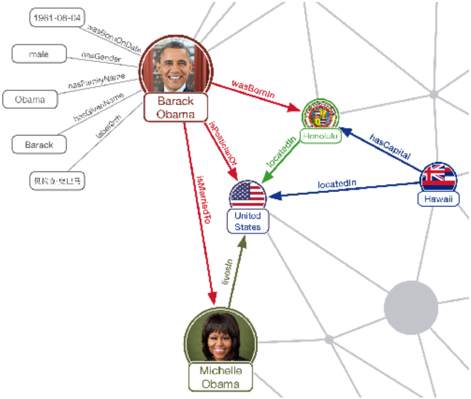


[www.laurentbrouat.com/the-social-network-of-jesus/]

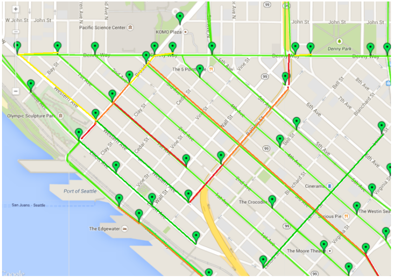
The Game of Thrones social network:



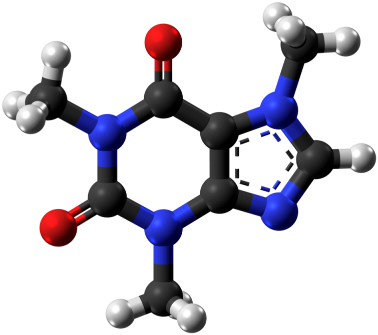
[Andrew Beveridge & Jie Shan (2016) Network of Thrones, Math Horizons `https://www.tandfonline.com/doi/abs/10.4169/mathhorizons.23.4.18`]



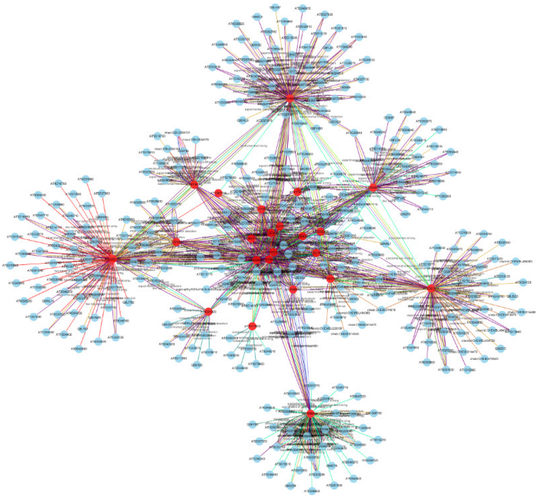
[Zhou, Zhixuan, et al. "Fake news detection via NLP is vulnerable to adversarial attacks." arXiv preprint 1901.09657 (2019).]



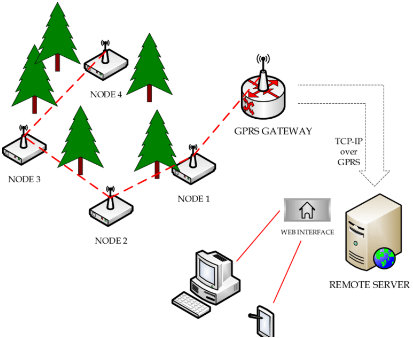
[https://www.flir.eu/discover/traffic/urban/real-time-traffic-information/]



[https://en.wikipedia.org/wiki/Caffeine]



[gmdd.shgmo.org/Computational-Biology/ANAP/ANAP_V1.1/help/anap-userguide/manual.html]



[L. Bencini et al. Wireless sensor networks for on-field agricultural management process. Wireless Sensor Networks: application-centric design (pp. 1-21). InTech.]

IMDB


The Godfather (1972) - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://uk.imdb.com/title/tt0068646/
Go
IMDb
Earth's Biggest Movie Database™
NOW PLAYING
MOVIE / TV NEWS
MY MOVIES
DVD / VIDEO
IMDb TV
MESSAGE BOARDS
SHOWTIMES & TICKETS
IMDbpro
IMDb Resume
Home | Top Movies | Photos | Independent Film | GameBase | Browse | Help
Login | Register
search All
go more | tips
IMDb > The Godfather (1972)
SHOP GODFATHER... amazon.com
All
DVD VHS CD
The Godfather (1972)
photo gallery
message boards
view trailer
add to My Movies
IMDb Pro details
Directed by
Francis Ford Coppola
Register or login to rate this title
User Rating: 9.1/10 (202,038 votes)
Top 250: #1
more
Writing credits
Mario Puzo (novel)
Mario Puzo (screenplay) ...
more
Quicklinks
main details
Top Links
- trailers
- full cast and crew
- trivia
- official sites
- memorable quotes
Overview
main details
- combined details
- full cast and crew
- company credits
- tv schedule
Promotional
- taglines
- trailers
- posters
- photo gallery
Awards & Reviews
- user comments
- external reviews
- newsgroup reviews
- awards
- user ratings
- recommendations
- message board
Plot & Quotes
- plot summary
- plot keywords
- Amazon.com
Photo Gallery (see all 52 photos)
more
Release Date: 24 March 1972 (USA) more
Genre: Crime / Drama more
Plot Outline: The aging patriarch of an organized crime dynasty transfers control of his clandestine empire to his reluctant son. more
Plot Keywords: No Opening Credits / Greed / Trilogy / Blood / Bloody Violence more
Awards: Won 3 Oscars. Another 19 wins & 17 nominations more
User Comments: The Godfather more
Cast (Cast overview, first billed only)
Marlon Brando ... Don Vito Corleone
Al Pacino ... Michael Corleone
James Caan ... Santino 'Sonny' Corleone
Richard S. Castellano ... Pete Clemenza (as Richard Castellano)
Continuing Teacher Education Courses
Earn the continuing education credits that you need for recertification, endorsements and pay increases... ONLINE!
ADMINISTRATION
TECHNOLOGY
SPECIAL EDUCATION
MORE
EARN 3 CTE CREDITS IN 3 WEEKS
Click to Learn More
University of Phoenix
advertisement
Done

IMDB Schema (simplified)

acts_in

actor	movie
1	1
2	2
3	3
3	4

directed_by

movie	director
1	1
2	2
3	2
4	3

Actor

id	gender
1	m
2	m
3	f

Movie

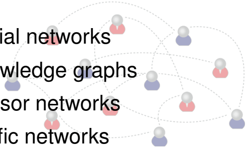
id	year	AA
1	1949	no
2	1982	no
3	1999	yes
4	2004	no

Director

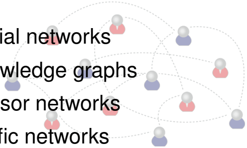
id	born
1	1911
2	1951
3	1962

Real-world networks

- Social networks
- Knowledge graphs
- Sensor networks
- Traffic networks
- . . .



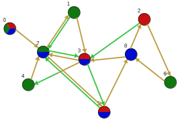
Real-world networks



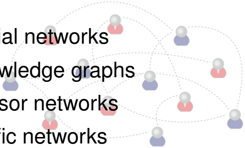
- Social networks
- Knowledge graphs
- Sensor networks
- Traffic networks
- . . .

Model

Abstract: graphs



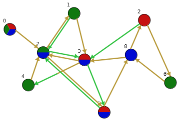
Real-world networks



- Social networks
- Knowledge graphs
- Sensor networks
- Traffic networks
- ...

Model

Abstract: graphs



Semantics

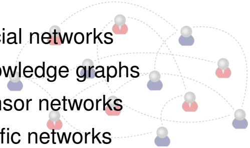
Predicate logic (relational)

$\forall x(r(x) \rightarrow \exists y(e(x,y) \wedge b(y)))$

$\exists z, x, y \neg(e(x,y) \wedge e(x,z) \wedge (e(y,z))$

...

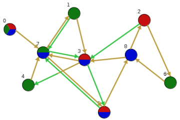
Real-world networks



- Social networks
- Knowledge graphs
- Sensor networks
- Traffic networks
- . . .

Model ↓

Abstract: graphs



Semantics ↑

Predicate logic (relational)

$\forall x(r(x) \rightarrow \exists y(e(x,y) \wedge b(y)))$

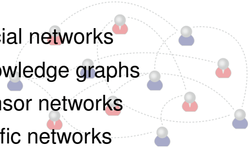
$\exists z, x, y \neg(e(x,y) \wedge e(x,z) \wedge (e(y,z))$

. . .

$+$*uncertainty* →

- Probabilistic Logic
- Statistical (Random) Graph Theory
- Statistical Relational Learning (Probabilistic Logic Learning)
- Graph Learning and Mining
- Graph Neural Networks

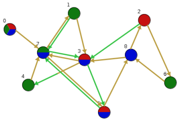
Real-world networks



- Social networks
- Knowledge graphs
- Sensor networks
- Traffic networks
- . . .

Model

Abstract: graphs



Semantics

Predicate logic (relational)

$\forall x (r(x) \rightarrow \exists y (e(x,y) \wedge b(y)))$

$\exists z, x, y \neg (e(x,y) \wedge e(x,z) \wedge (e(y,z))$

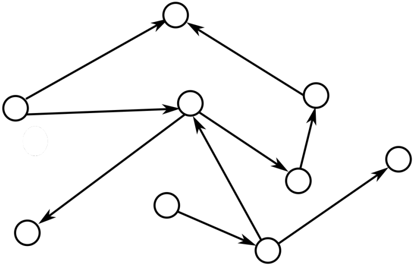
. . .

$+$*uncertainty*

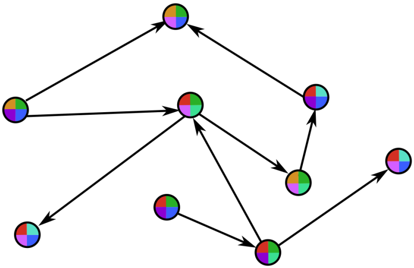
- Probabilistic Logic
- Statistical (Random) Graph Theory
- Statistical Relational Learning (Probabilistic Logic Learning)
- Graph Learning and Mining
- Graph Neural Networks

- Graph representations
- Learning and reasoning with graphs
- GNN approaches
- SRL approaches
- Comparisons & Integration

Graph Representations

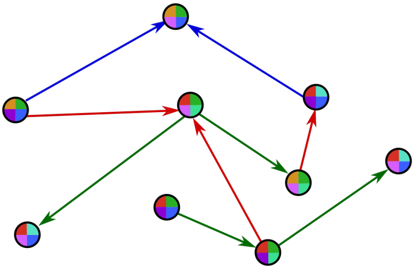


Graph: (V, E)



Graph: (V, E)

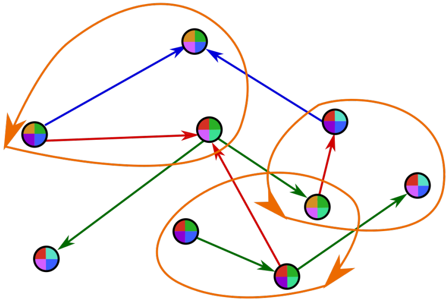
Attributed graph: $(V, E, \boldsymbol{A})$. Node attributes $\boldsymbol{A}$: *Boolean*, *categorical*, or *numeric*



Graph: (V, E)

Attributed graph: $(V, E, \boldsymbol{A})$. Node attributes $\boldsymbol{A}$: *Boolean*, *categorical*, or *numeric*

Attributed multirelational graph: $(V, \boldsymbol{E}, \boldsymbol{A})$. $\boldsymbol{E}$: set of different edge relations

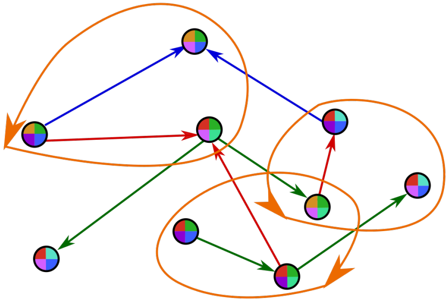


Graph: (V, E)

Attributed graph: $(V, E, \boldsymbol{A})$. Node attributes $\boldsymbol{A}$: *Boolean*, *categorical*, or *numeric*

Attributed multirelational graph: $(V, \boldsymbol{E}, \boldsymbol{A})$. $\boldsymbol{E}$: set of different edge relations

Attributed multirelational hyper-graph: $(V, \boldsymbol{R})$. $\boldsymbol{R}$: set of 1,2,3,...-ary relations (subsumes $\boldsymbol{A}$, $\boldsymbol{E}$)



Graph: (V, E)

Attributed graph: $(V, E, \boldsymbol{A})$. Node attributes $\boldsymbol{A}$: *Boolean*, *categorical*, or *numeric*

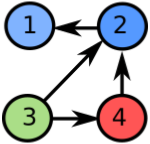
Attributed multirelational graph: $(V, \boldsymbol{E}, \boldsymbol{A})$. $\boldsymbol{E}$: set of different edge relations

Attributed multirelational hyper-graph: $(V, \boldsymbol{R})$. $\boldsymbol{R}$: set of 1,2,3,...-ary relations (subsumes $\boldsymbol{A}, \boldsymbol{E}$)

Examples for higher arity relations:

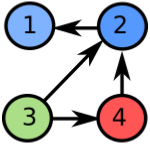
3-ary traffic network relation: *on_shortest_path*(location,location,location)
3-ary movie data: *made_contract*(agent,actor,movie)

Node	color
1	blue
2	blue
3	green
4	red

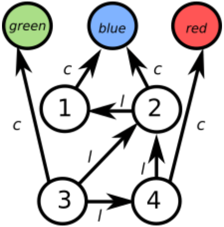


(a)

Node	c_blue	c_green	c_red
1	1	0	0
2	1	0	0
3	0	1	0
4	0	0	1



(b)

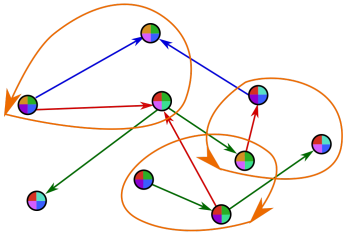


(c)

(a): unary, categorical values

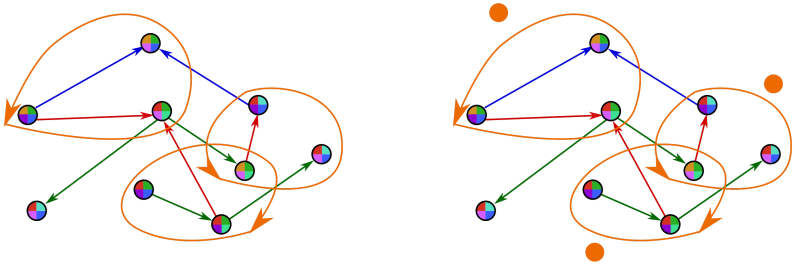
(b): unary, Boolean/binary values (one-hot encoding)

(c): binary relation between objects and attribute values materialized as nodes



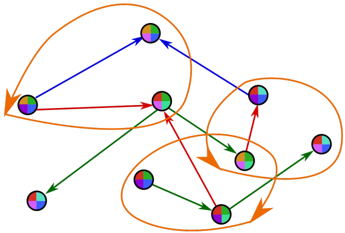
(a)

(a): as tuples of nodes

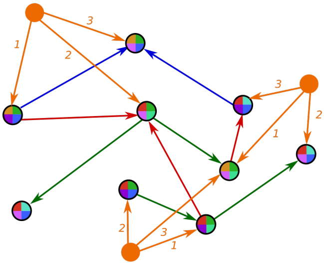


(a) (b)

- (a): as tuples of nodes
- (b): materialize tuples of nodes;



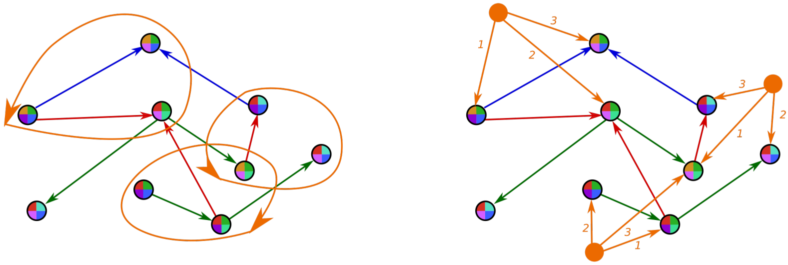
(a)



(b)

(a): as tuples of nodes

(b): materialize tuples of nodes;
connect tuple-nodes with entity-nodes by binary relations



(a) (b)

(a): as tuples of nodes

(b): materialize tuples of nodes;
connect tuple-nodes with entity-nodes by binary relations

➥ Boolean values and binary relations are enough in principle (but can be user-unfriendly).

n	number of nodes/vertices
$\mathcal{R}$	a *signature* of 1,2,3,...-ary relation symbols
$\boldsymbol{R}$	specific values of the relations in $\mathcal{R}$ in a graph $G = (V, \boldsymbol{R})$.
$G = (V, \boldsymbol{R})$	a graph with node set V, and relations $\boldsymbol{R}$
$\mathcal{G}(V, \mathcal{R})$	set of all graphs with node set V, and relations in the signature $\mathcal{R}$
$\Delta\mathcal{G}(V, \mathcal{R})$	set of all probability distributions over $\mathcal{G}(V, \mathcal{R})$

Generally assume that $V = \{1, \ldots, n\}$, and $i, j, \ldots \in \mathbb{N}$ denote nodes.

Arrays, Matrices, Tensors

Attributes:

n-dimensional array (entries categorical);

Color:

$$(\textit{blue}, \textit{green}, \ldots, \textit{red}, \textit{blue})$$

k different attributes collected in $n \times k$-dimensional matrix

Binary relations:

$(n \times n)$-dimensional *adjacency matrix*;

Friends:

$$\begin{pmatrix} 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 1 \end{pmatrix}$$

l different relations may be collected in a $(n \times n \times l)$-dimensional *adjacency tensor*

Arrays, Matrices, Tensors

Attributes:

n-dimensional array (entries categorical);

Color:

$$(\textit{blue}, \textit{green}, \ldots, \textit{red}, \textit{blue})$$

k different attributes collected in $n \times k$-dimensional matrix

Binary relations:

$(n \times n)$-dimensional *adjacency matrix*;

Friends:

$$\begin{pmatrix} 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 1 \end{pmatrix}$$

l different relations may be collected in a $(n \times n \times l)$-dimensional *adjacency tensor*

Logic

(Boolean) relations of all arities:

List positive and negative facts as a *knowledge base* of (negated) *ground atoms* :

$$\begin{aligned} \textit{c\_blue}(17) \\ \neg \textit{c\_red}(21) \\ \textit{friends}(2{,}53) \\ \neg \textit{friends}(53{,}8) \\ \vdots \end{aligned}$$

Arrays, Matrices, Tensors

Attributes:

n-dimensional array (entries categorical);

Color:

$$(\textit{blue}, \textit{green}, \ldots, \textit{red}, \textit{blue})$$

k different attributes collected in $n \times k$-dimensional matrix

Binary relations:

$(n \times n)$-dimensional *adjacency matrix*;

Friends:

$$\begin{pmatrix} 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 1 \end{pmatrix}$$

l different relations may be collected in a $(n \times n \times l)$-dimensional *adjacency tensor*

Logic

(Boolean) relations of all arities:

List positive and negative facts as a *knowledge base* of (negated) *ground atoms* :

$$\begin{aligned} \textit{c\_blue}(17) \\ \neg \textit{c\_red}(21) \\ \textit{friends}(2{,}53) \\ \neg \textit{friends}(53{,}8) \\ \vdots \end{aligned}$$

Knowledge Graphs

Triplet notation

$$(\textit{Barack Oboma}, \textit{was\_born\_in}, \textit{Honolulu})$$

can be seen as “infix” notation for ground atoms:

$$\textit{was\_born\_in}(\textit{Barack Oboma}, \textit{Honolulu})$$

Pros and Cons

Tensors:
- \+ compact, computationally efficient
- \+ support for categorical (attribute) values
- \- incomplete data requires '?' or 'NaN' values

Logic:
- \+ coherent treatment of relations of all arities
- \+ uniform treatment of positive and negative facts
- \+ natural support for incomplete knowledge
- \- non binary relations require special treatment
- \- non-compact

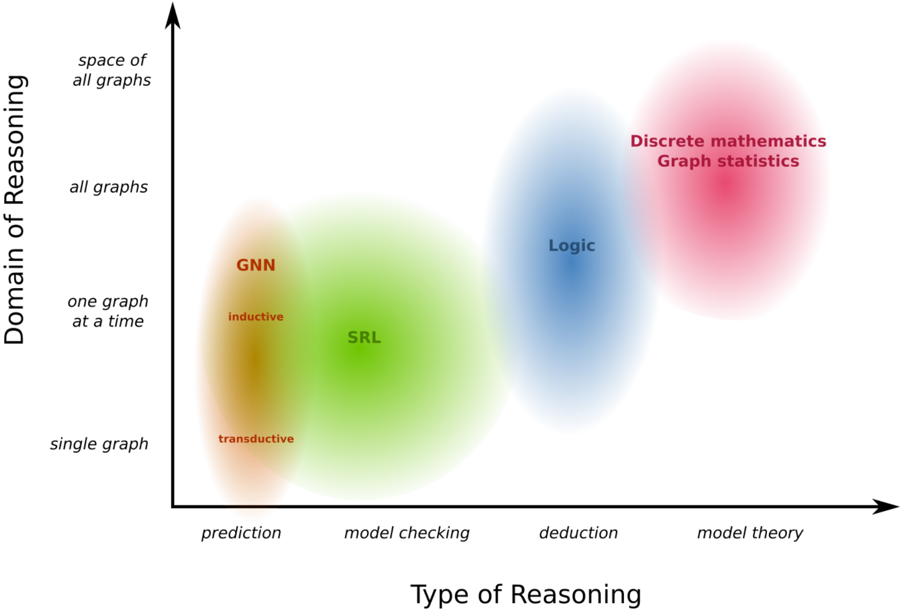
Missing information

Tensors: '0' in adjacency matrix can mean: “there is no edge”, or “there is no confirmed edge”

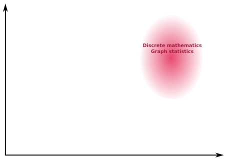
Logic: Positive and negative information are equally definitive. Missing information visible as incomplete knowledge base

To handle semantic asymmetry: *closed-world assumption* used to postulate that all facts not contained in the knowledge base are assumed to be *false* by default.

Reasoning



Domain of Reasoning
space of all graphs
all graphs
one graph at a time
single graph
GNN
inductive
transductive
SRL
Logic
Discrete mathematics
Graph statistics
prediction
model checking
deduction
model theory
Type of Reasoning

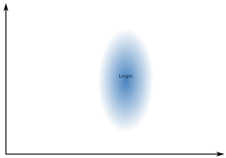


Given: a probabilistic model for the random generation/evolution of graphs.

Question: what is the probability that the graph becomes (stays) connected, as the number of nodes goes to infinity?

➥Or many other questions about the global properties of a random graph model.

➥Mostly (human powered) mathematics, rather than algorithmic reasoning



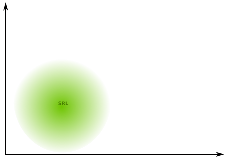
Given: a knowledge base

$$\forall x \exists y \; follows(x, y)$$
$$\exists y \neg \exists x \; follows(x, y)$$

Question: Does the knowledge base imply a given query statement?

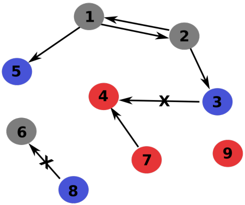
$$(\exists y \exists^{\geq 2} x \; follows(x, y)) \vee \exists^{\geq 10.000} x \; ?$$

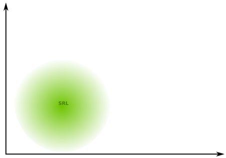
➥Algorithmic reasoning implemented by *theorem provers*.



Given: a generative probabilistic model for graphs.

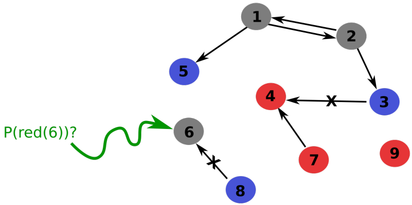
Question: for a single partially observed graph, what are the probabilties of unobserved features?

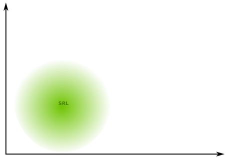




Given: a generative probabilistic model for graphs.

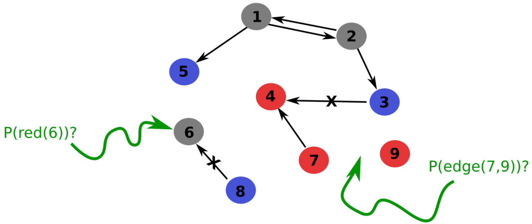
Question: for a single partially observed graph, what are the probabilties of unobserved features?

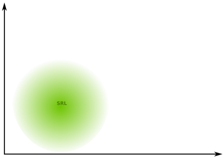




Given: a generative probabilistic model for graphs.

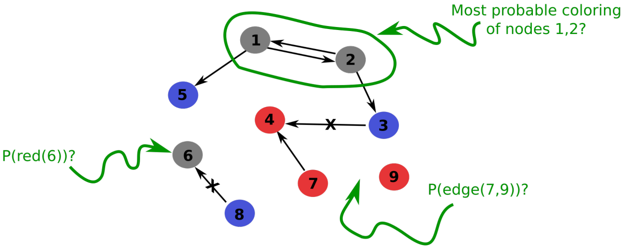
Question: for a single partially observed graph, what are the probabilties of unobserved features?

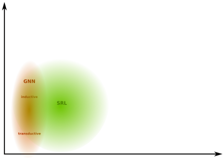




Given: a generative probabilistic model for graphs.

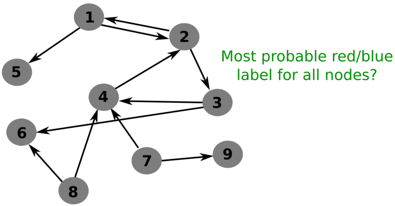
Question: for a single partially observed graph, what are the probabilties of unobserved features?





Given: a descriminative model for specific node label.

Question: for an input graph (edges, node attributes), what are predicted node labels?

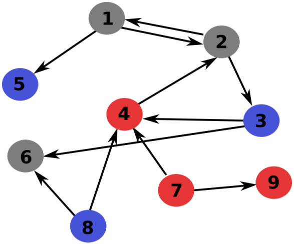


➥Similarly: link prediction, graph classification.

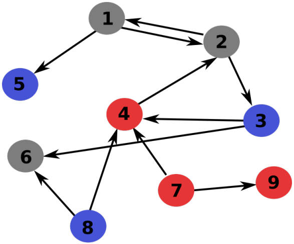
Learning

Learning and Reasoning about a single graph:

Training data

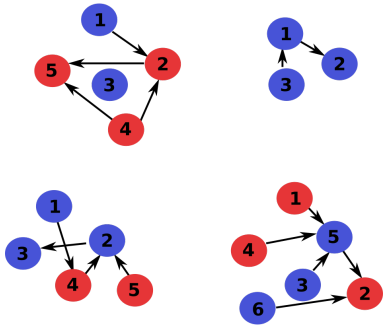


Reasoning domain



Learning and Reasoning about different graphs:

Training data



Reasoning domains (a.k.a. test cases)