Relational inductive biases, deep learning, and graph networks

Battaglia et al.

https://arxiv.org/abs/1806.01261

Robert Różański

ML Gdańsk 29.10.2018

Deep Learning challenges

complex language and scene understanding

reasoning about structured data

transfer learning beyond the learning conditions

learning from small amount of experience

unstructured approaches

weak assumptions

low inductive bias

high data requirements

high computation requirements

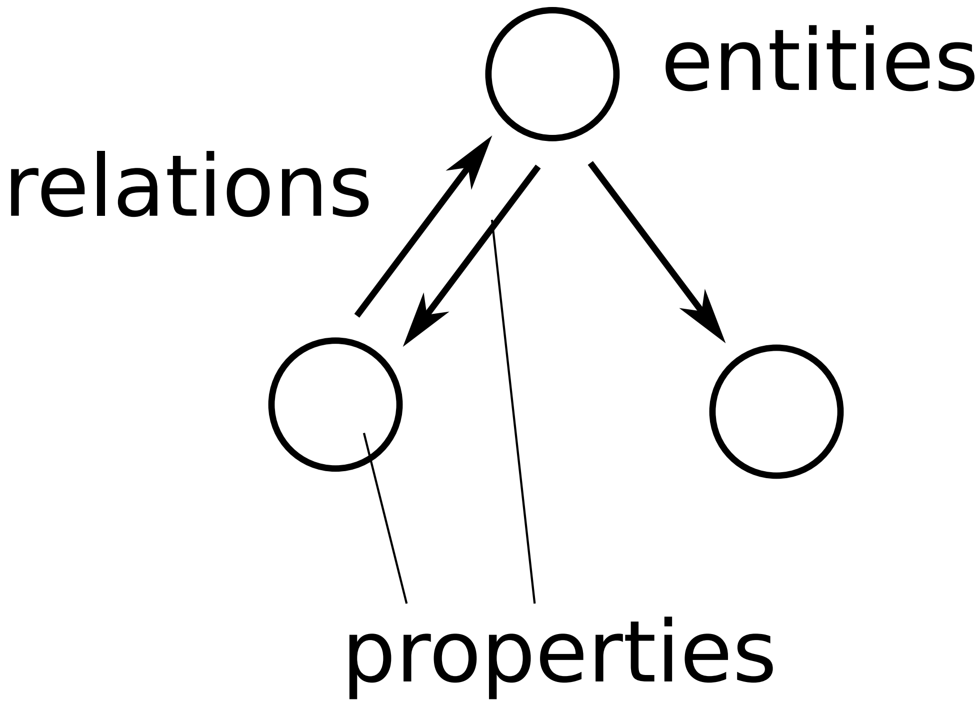
transfer learning (limited)

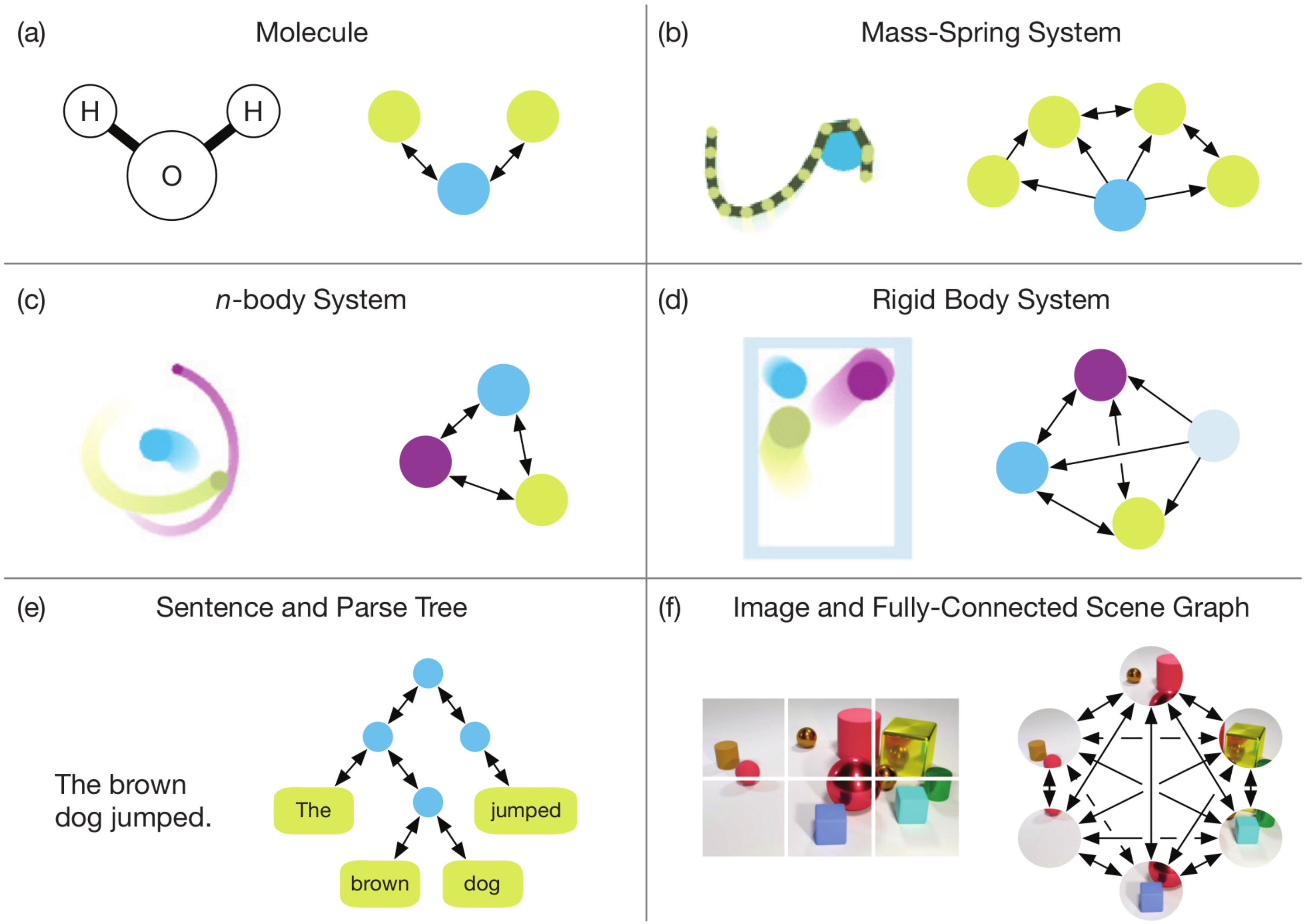
structured approaches	unstructured approaches
strong a priori assumptions about data structures and computation	weak assumptions
high inductive bias	low inductive bias
low data requirements	high data requirements
low computation requirements	high computation requirements
combinatorial generalisation	transfer learning (limited)

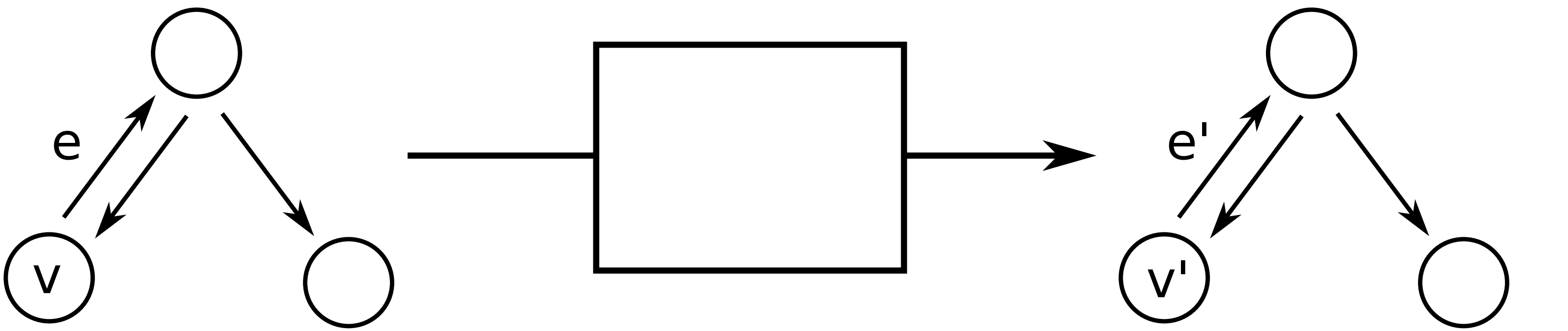
structured approaches	unstructured approaches
strong a priori assumptions about data structures and computation	weak assumptions
high inductive bias	low inductive bias
low data requirements	high data requirements
low computation requirements	high computation requirements
combinatorial generalisation	transfer learning (limited)

Which structures to choose?

Which structures to choose?

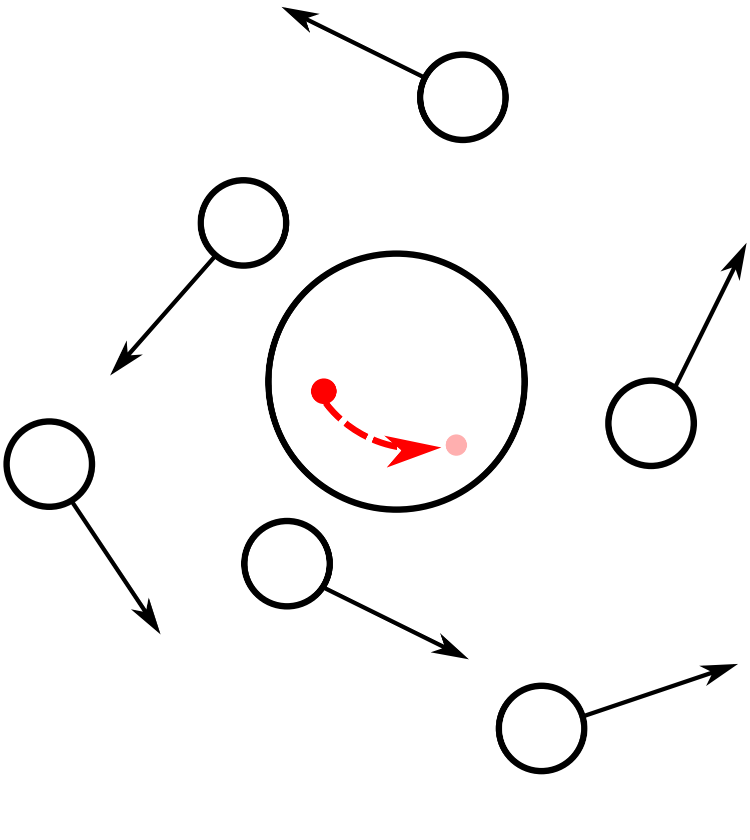



(a)
Molecule
H
O
H
(b)
Mass-Spring System
(c)
n-body System
(d)
Rigid Body System
(e)
Sentence and Parse Tree
The brown
dog jumped.
The
jumped
brown
dog
(f)
Image and Fully-Connected Scene Graph

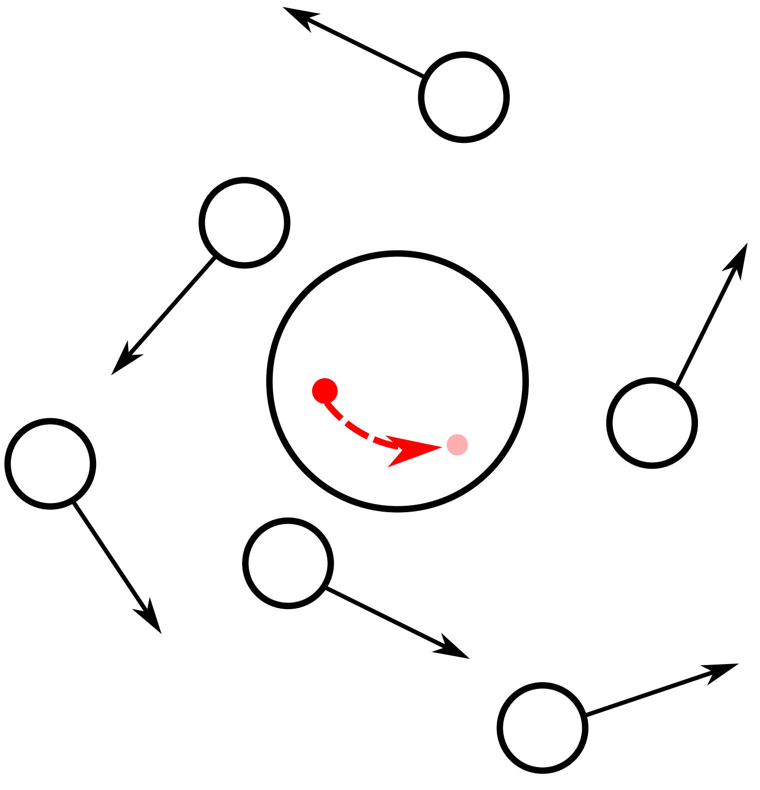


e
v
e'
v'

order invariance: predicting center of mass

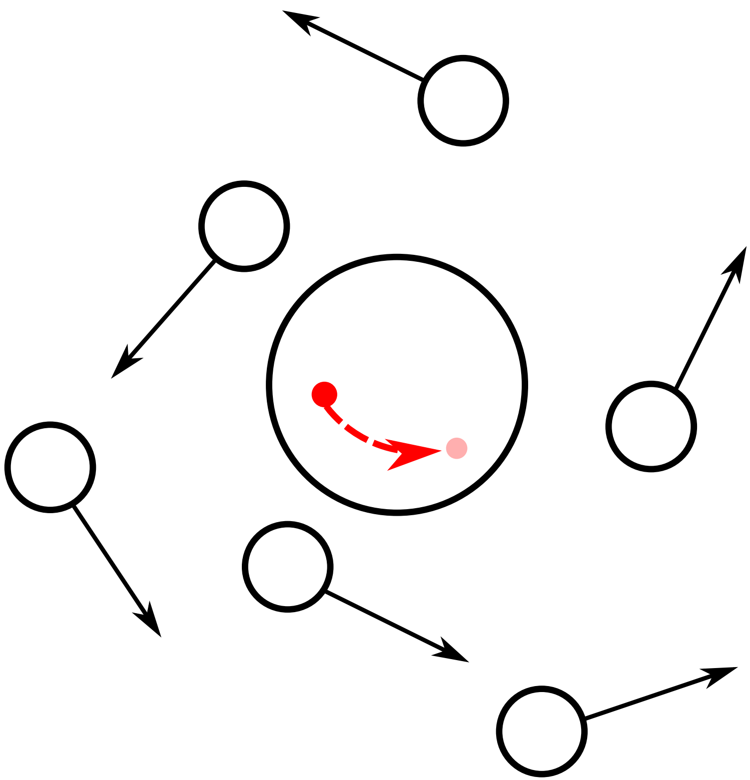


order invariance: predicting center of mass



e.g. for MLP order matters
(n! permutations)

order invariance: predicting center of mass



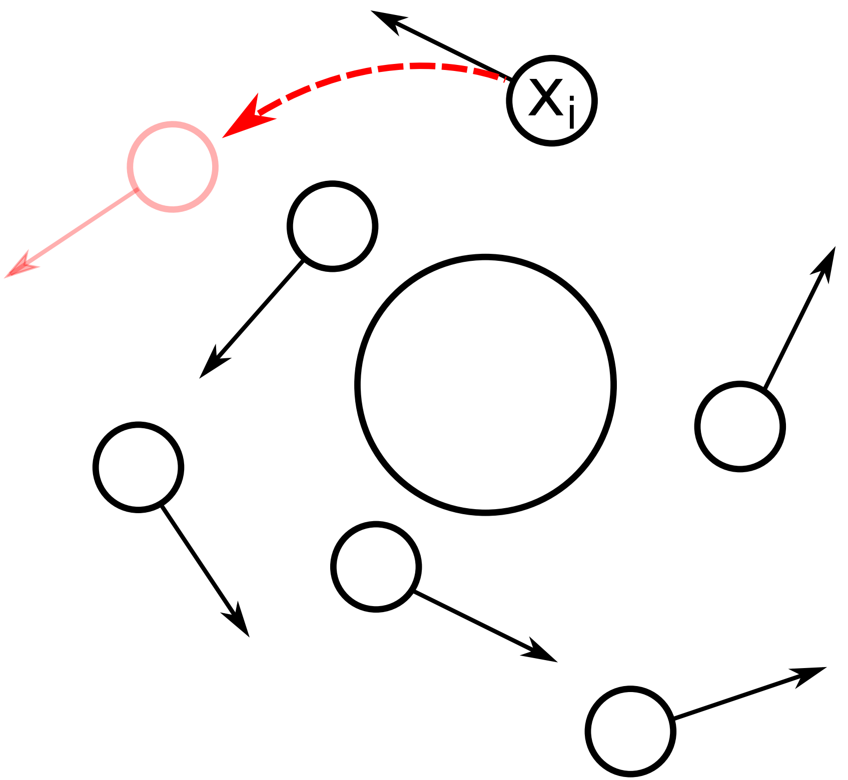
e.g. for MLP order matters
(n! permutations)

Forcing invariance:

1. compute features (per planet)

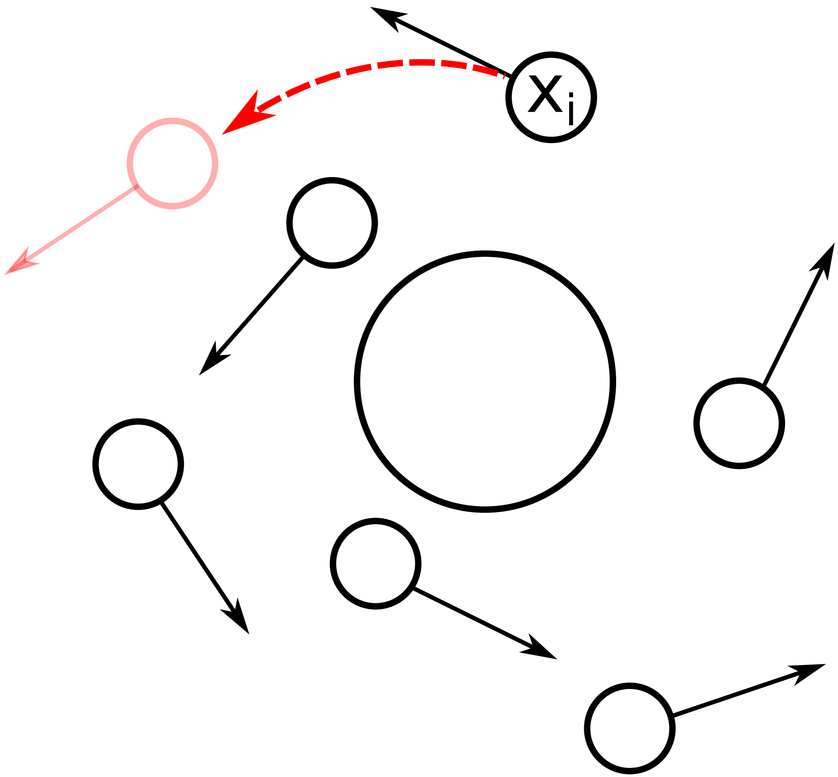
2. aggregate using order-invariant function

pairwise relations: predicting planet's position



planet's future position depends on parameters of all other planets:

pairwise relations:
predicting planet's position

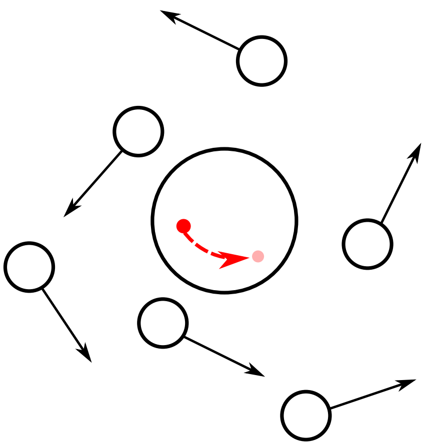


planet's future position depends on parameters of all other planets:

$$\mathbf{x}'_i = f(\mathbf{x}_i, \sum_j g(\mathbf{x}_i, \mathbf{x}_j))$$

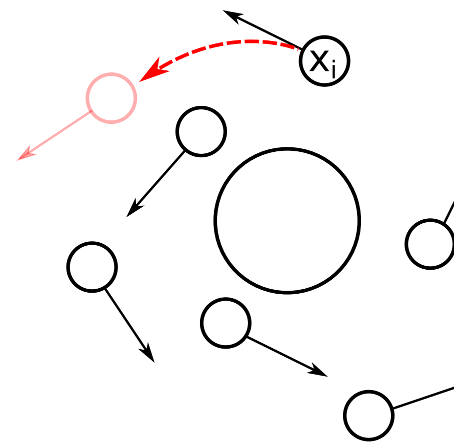

no connections
fully connected
x_i

no connections



fully connected

$$\mathbf{x}'_i = f(\mathbf{x}_i, \textstyle\sum_j g(\mathbf{x}_i, \mathbf{x}_j))$$



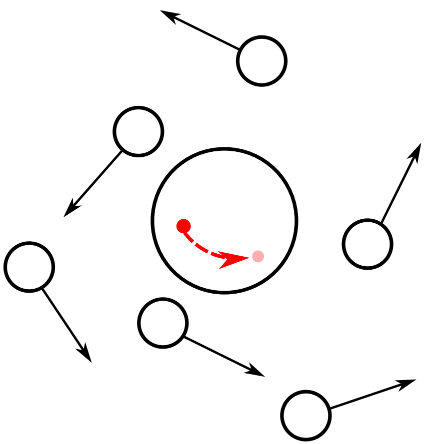
no connections

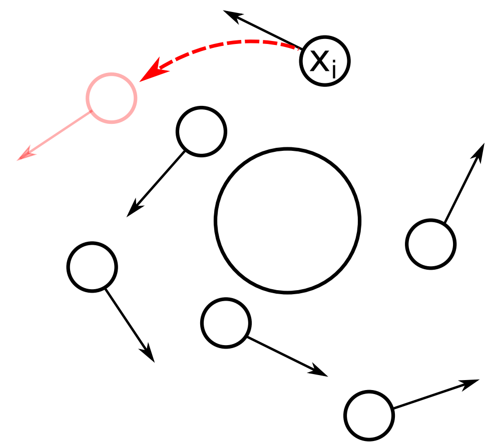
intermediate

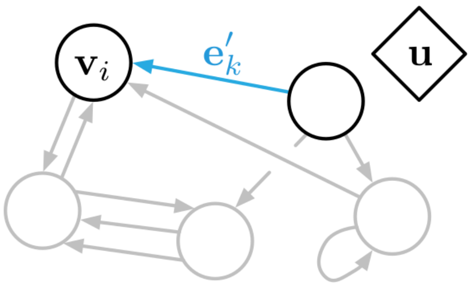
$$x_i' = f(\mathbf{x}_i, \sum_{j \in \delta(i)} g(\mathbf{x}_i, \mathbf{x}_j))$$

fully connected

$$\mathbf{x}_i' = f(\mathbf{x}_i, \sum_j g(\mathbf{x}_i, \mathbf{x}_j))$$

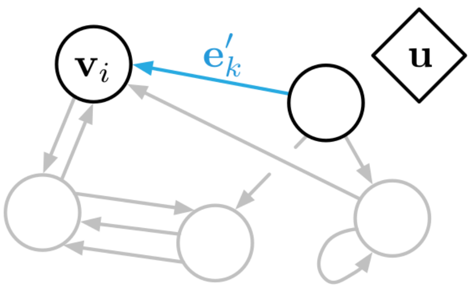






(a) Edge update

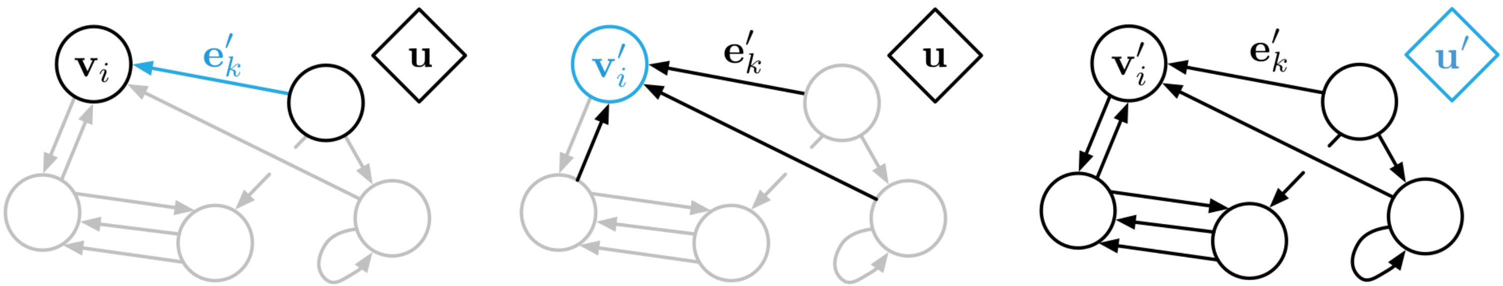
$$\mathbf{e}'_k = \phi^e\left(\mathbf{e}_k, \mathbf{v}_{r_k}, \mathbf{v}_{s_k}, \mathbf{u}\right)$$



(a) Edge update (b) Node update

$$\mathbf{e}'_k = \phi^e\left(\mathbf{e}_k, \mathbf{v}_{r_k}, \mathbf{v}_{s_k}, \mathbf{u}\right)$$
$$\mathbf{v}'_i = \phi^v\left(\bar{\mathbf{e}}'_i, \mathbf{v}_i, \mathbf{u}\right)$$

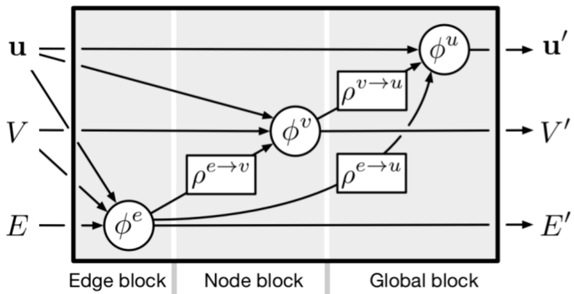
$$\bar{\mathbf{e}}'_i = \rho^{e\rightarrow v}\left(E'_i\right)$$



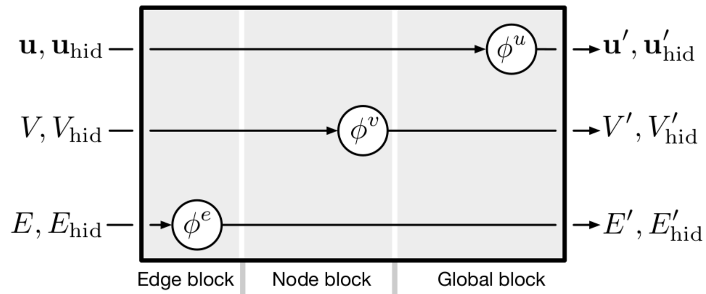
(a) Edge update (b) Node update (c) Global update

$$\begin{aligned}
\mathbf{e}'_k &= \phi^e\left(\mathbf{e}_k, \mathbf{v}_{r_k}, \mathbf{v}_{s_k}, \mathbf{u}\right) \\
\mathbf{v}'_i &= \phi^v\left(\bar{\mathbf{e}}'_i, \mathbf{v}_i, \mathbf{u}\right) \\
\mathbf{u}' &= \phi^u\left(\bar{\mathbf{e}}', \bar{\mathbf{v}}', \mathbf{u}\right)
\end{aligned}$$

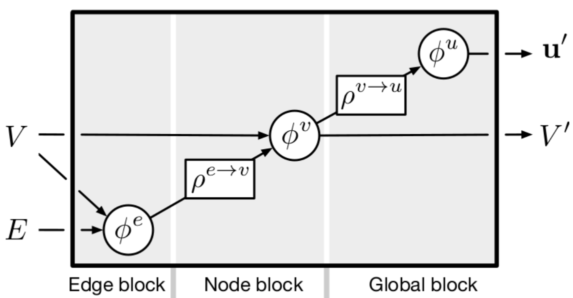
$$\begin{aligned}
\bar{\mathbf{e}}'_i &= \rho^{e\rightarrow v}\left(E'_i\right) \\
\bar{\mathbf{e}}' &= \rho^{e\rightarrow u}\left(E'\right) \\
\bar{\mathbf{v}}' &= \rho^{v\rightarrow u}\left(V'\right)
\end{aligned}$$



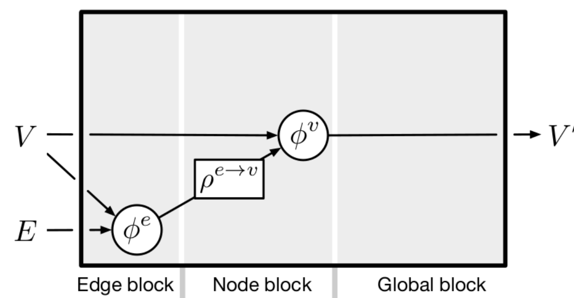
(a) Full GN block



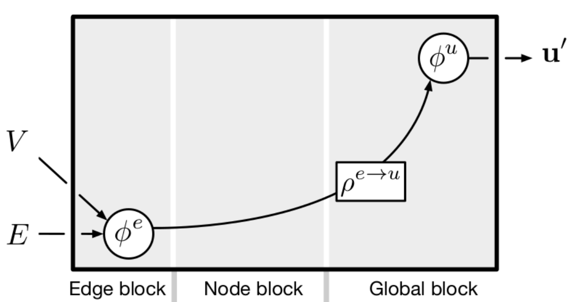
(b) Independent recurrent block



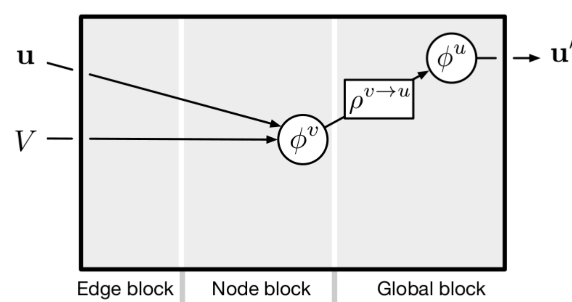
(c) Message-passing neural network



(d) Non-local neural network



(e) Relation network



(f) Deep set