Relational Inductive Biases, Deep Learning, & Graph Networks

Peter Baggalia et al. 2018

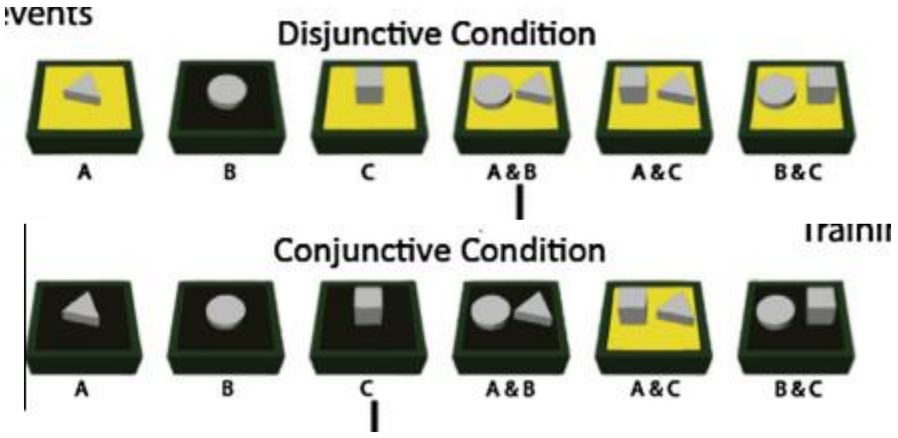
Presenter: Robert Thorstad

Motivation

Two Features of Human Learning

Inductive Biases

"Events have one cause"



Lucas, Bridgers, Griffiths, & Gopnik, 2014

Compositional

"Will it fall?"

Battaglia, Hamrick, & Tenenbaum, 2013

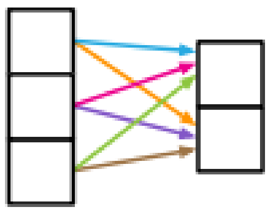
Question: what kinds of inductive biases exist in current models, and are they general enough?

Question: if I know if a tower of 3 blocks will stand, I also know this for: more blocks, tower of cups, large rocks, etc….does a model?

1. Inductive Biases

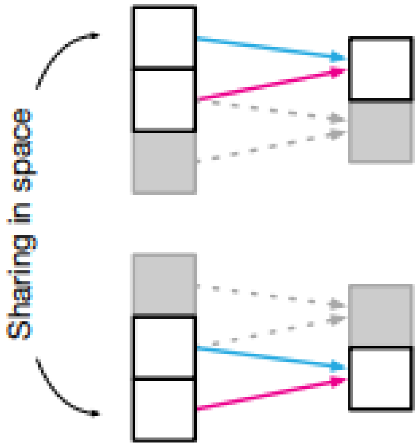
Existing models have only certain kinds of inductive biases

Little bias



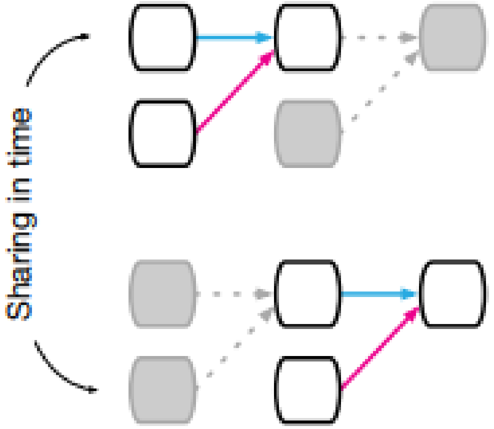
(a) Fully connected

Spatially Connected



(b) Convolutional

Sequential



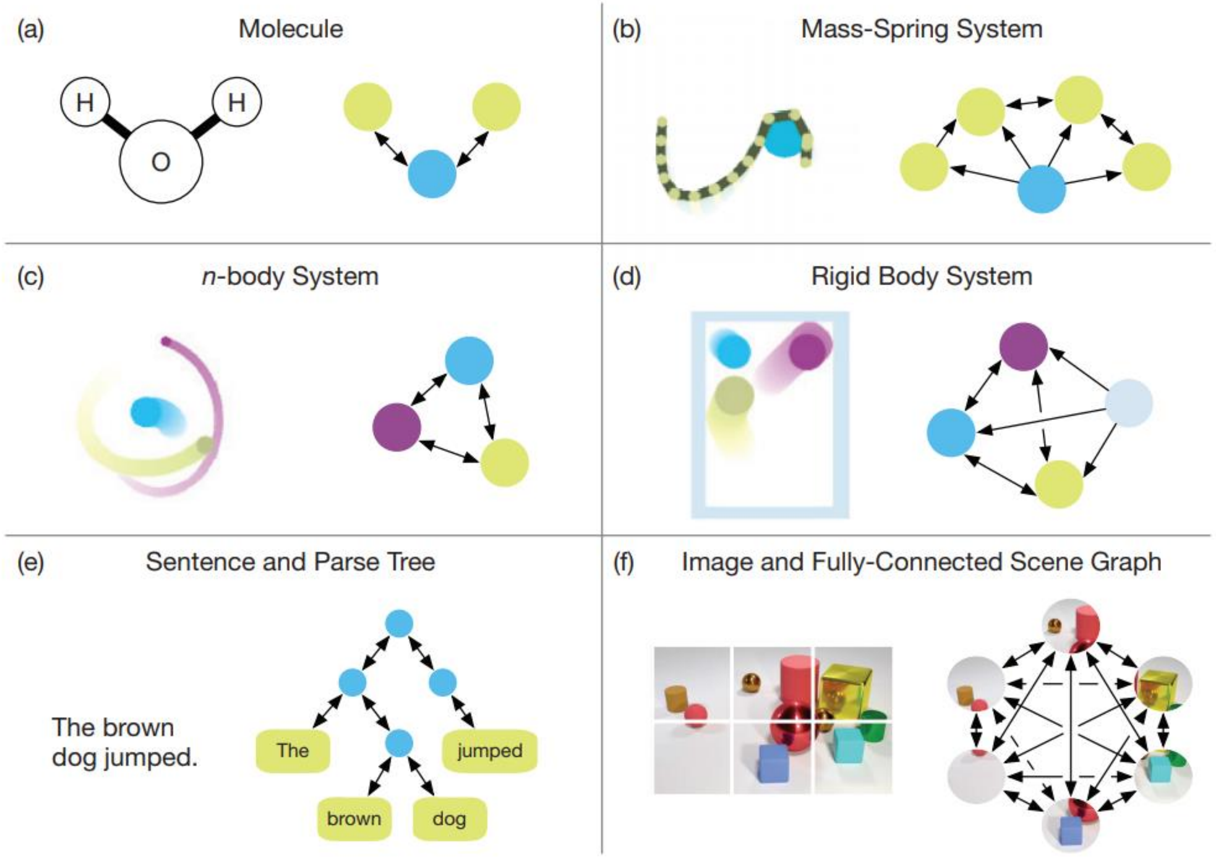
(c) Recurrent

Also:

L1, L2
Etc.

1. Inductive Biases

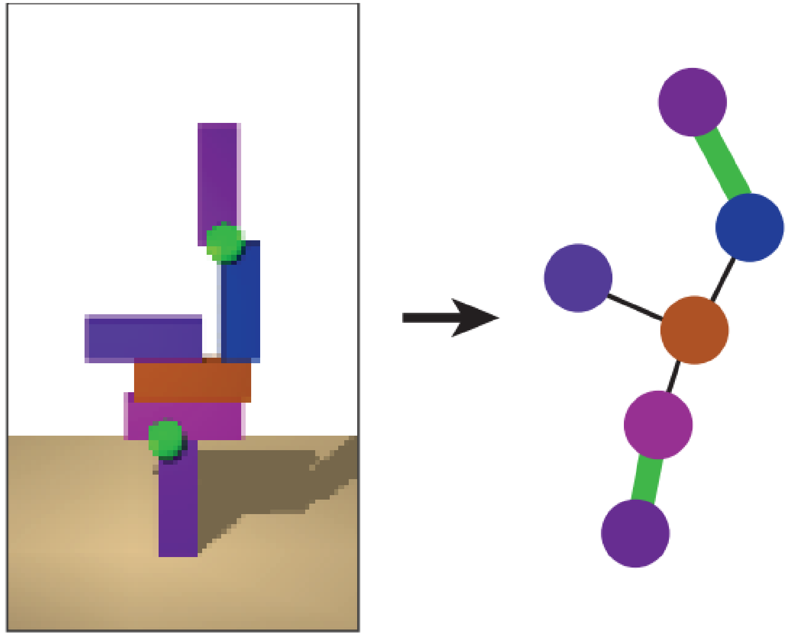
But can we model data that looks like…?



1. Inductive Biases

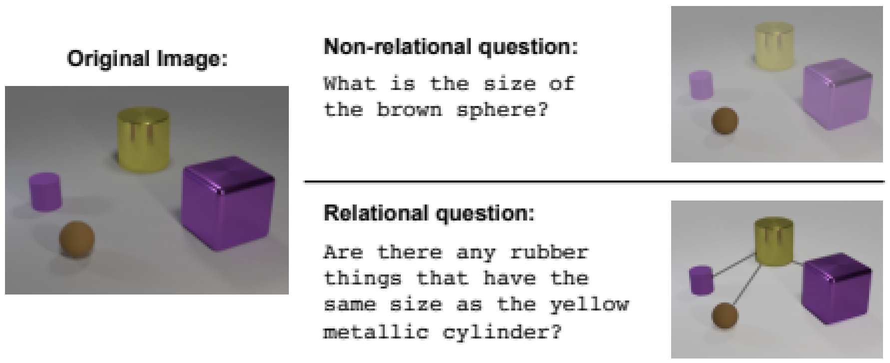
For example, how would we model these problems?

Glue to make tower stable



Hamrick et al, 2018

Is there an X like Y?



ure 1: **An illustrative example from the CLEVR dataset of relational reasoning**. An image ıtaining four objects is shown alongside non-relational and relational questions. The relational stion requires explicit reasoning about the relations between the four objects in the image, whereas non-relational question requires reasoning about the attributes of a particular object.

Santoro, 2017

1. Inductive Biases

We would like a model that can accommodate any ARBITRARY inductive bias

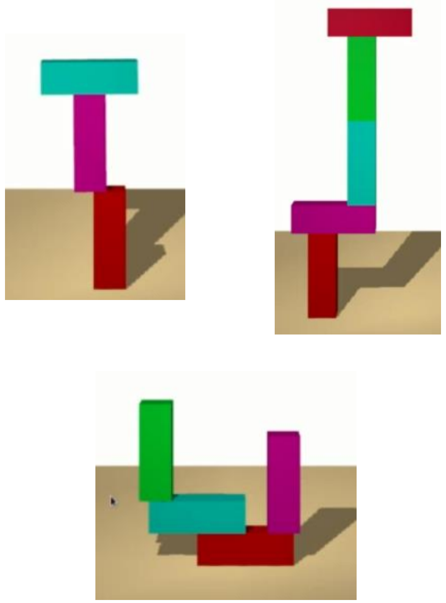
Component	Entities	Relations	Rel. inductive bias	Invariance
Fully connected	Units	All-to-all	Weak	-
Convolutional	Grid elements	Local	Locality	Spatial translation
Recurrent	Timesteps	Sequential	Sequentiality	Time translation
Graph network	Nodes	Edges	Arbitrary	Node, edge permutations

Table 1: Various relational inductive biases in standard deep learning components. See also Section 2.

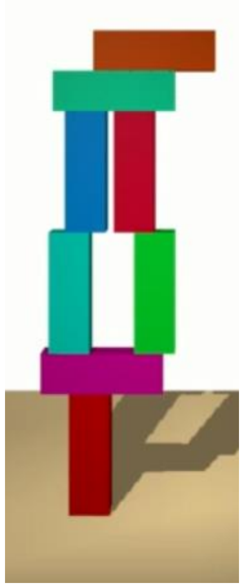
2. Generalization

People can generalize beyond training conditions:

Learn these:



Generalize to:



Also:

A world with less gravity

Blocks with different material

Blocks of different size

…

Hamrick et al, 2018

2. Generalization

But can a neural network generalize in a similar way?

Train

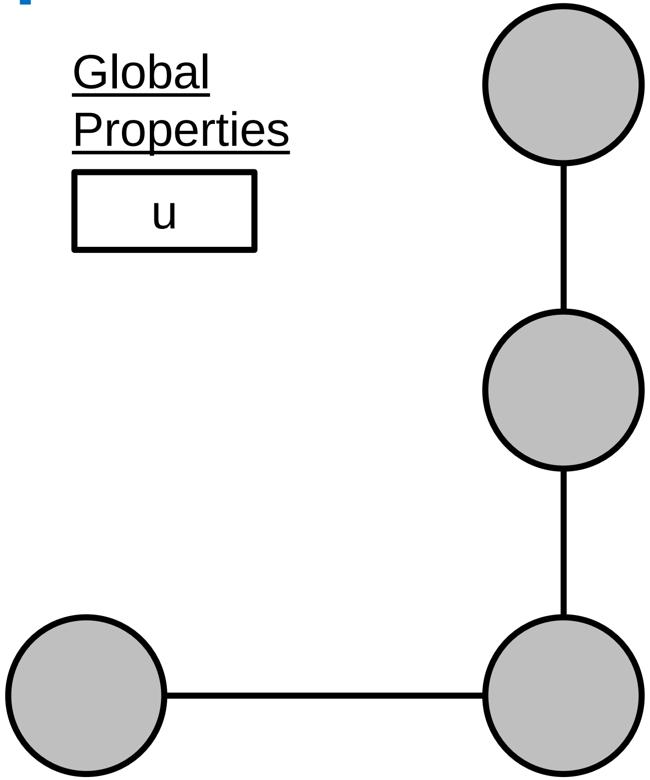
Test

E.g.: longer sentences, deeper parse tree, towers with more blocks…

Graph Neural Networks

Graph Neural Network

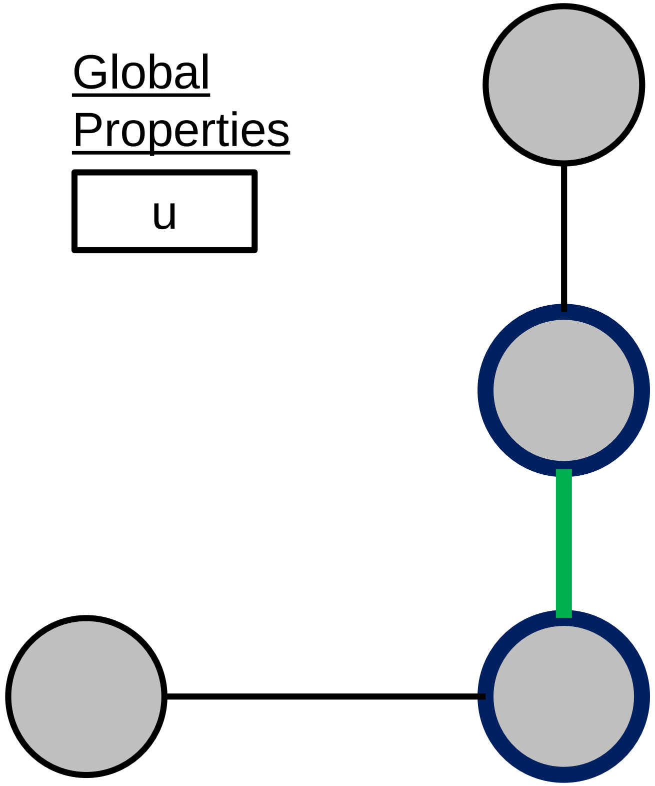
Neural network over ANY graph structure

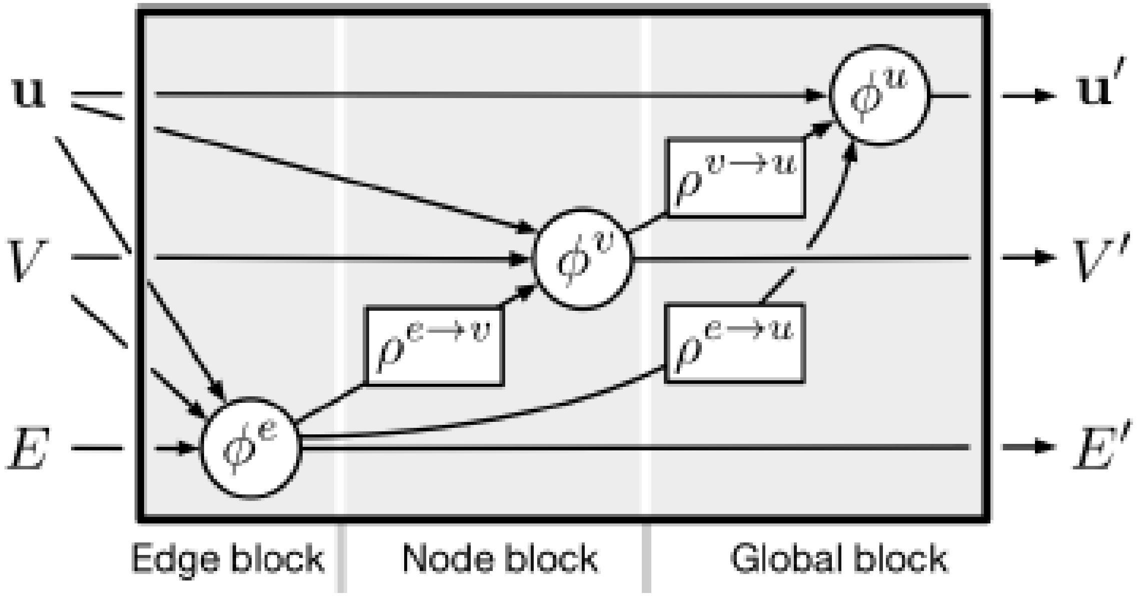


A graph is specified by:

Training: A Graph Network (GN) Block

Update an edge

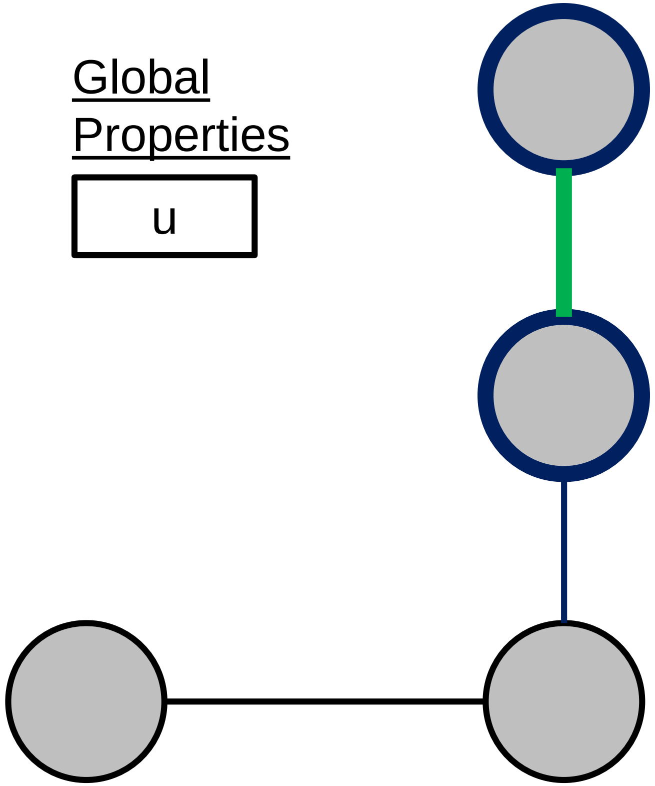


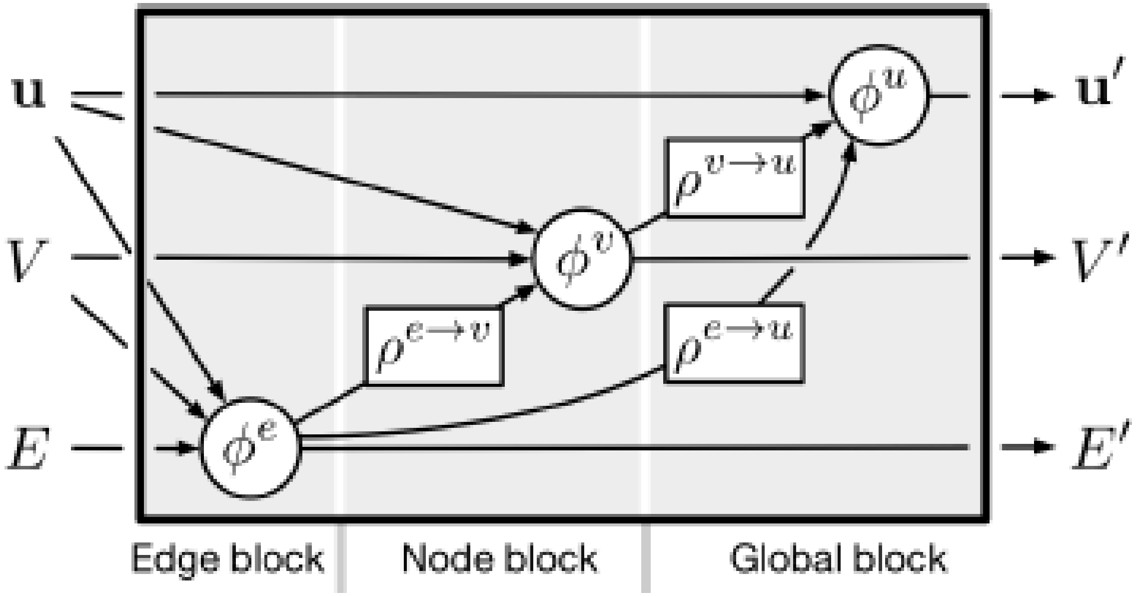


(a) Full GN block

Training: A Graph Network (GN) Block

Updating another edge involves same (shared) weights
In principle, could thus add nodes/edges after training

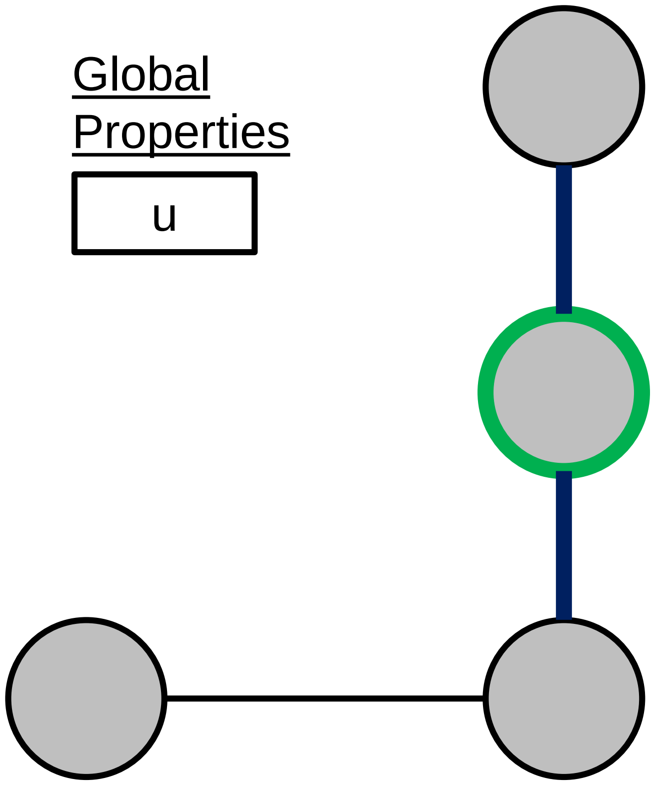


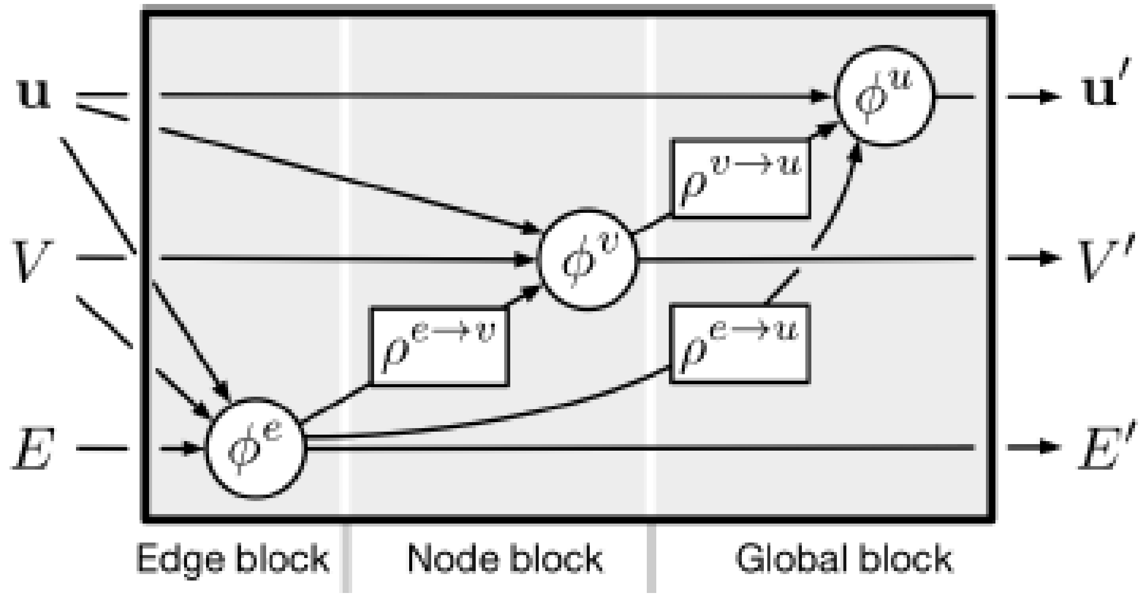


(a) Full GN block

Training: A Graph Network (GN) Block

Update a node

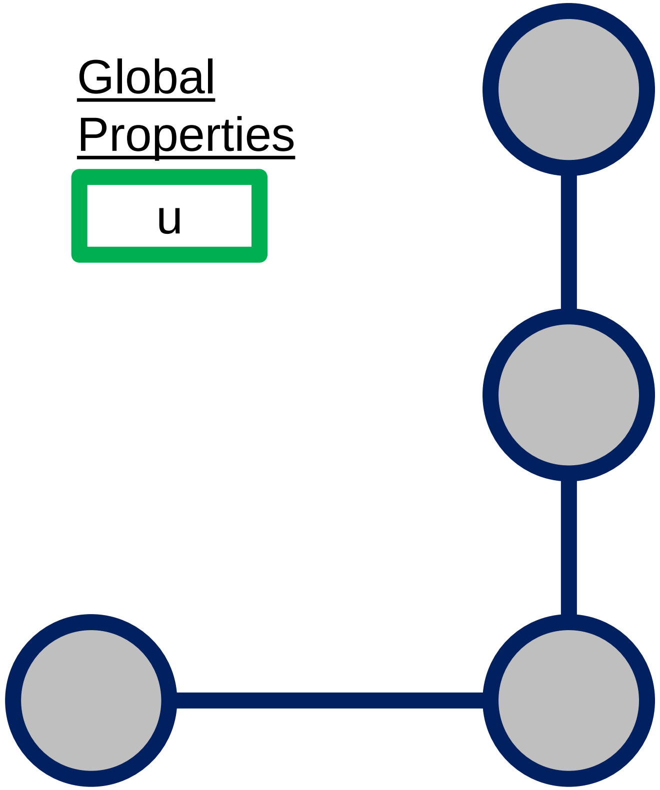


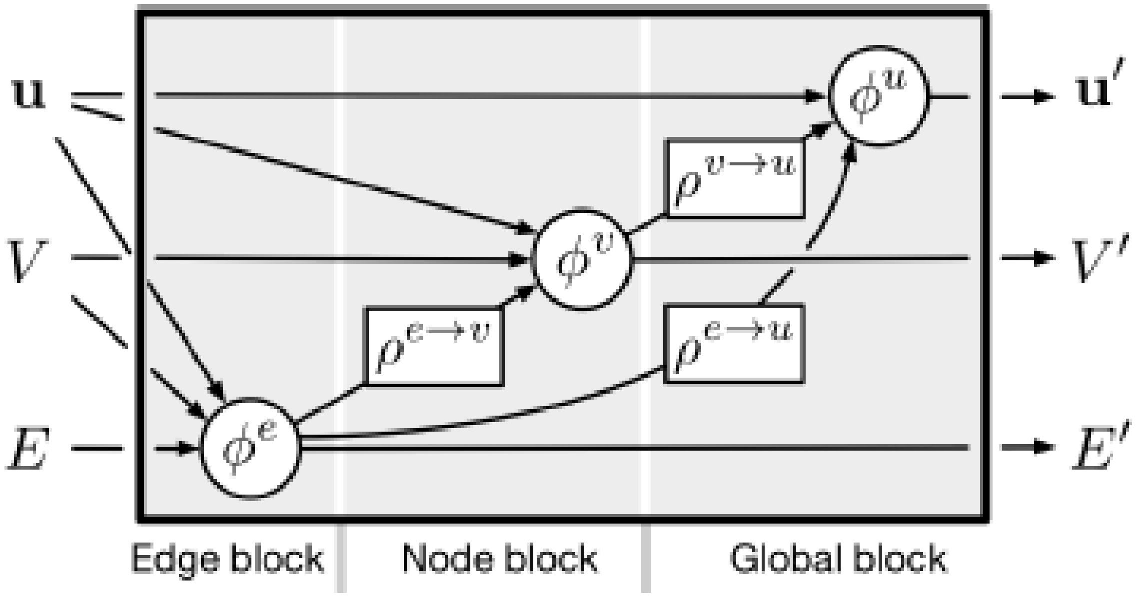


(a) Full GN block

Training: A Graph Network (GN) Block

Update global attributes



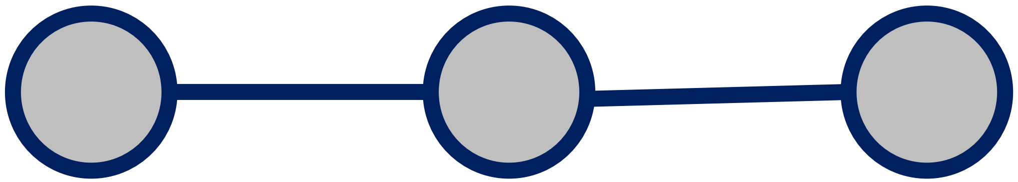


(a) Full GN block

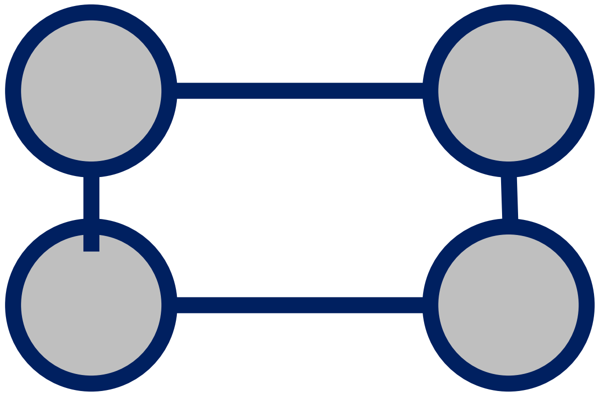
Idea: GNs have Arbitrary Inductive Bias

Can learn these graphs

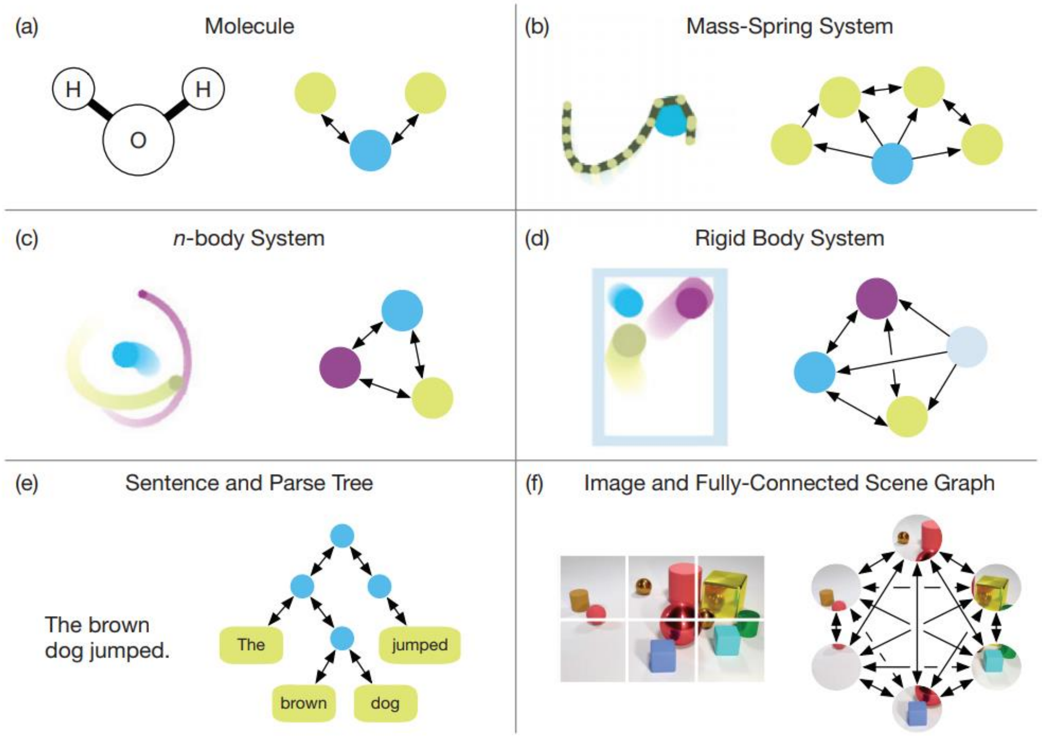
e.g. a sentence



e.g. an image



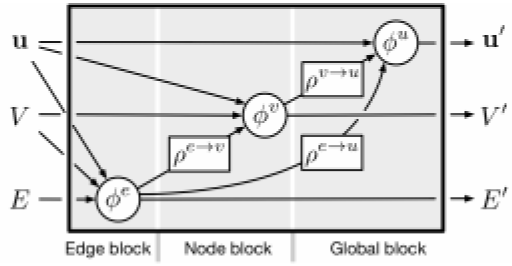
But also other structures



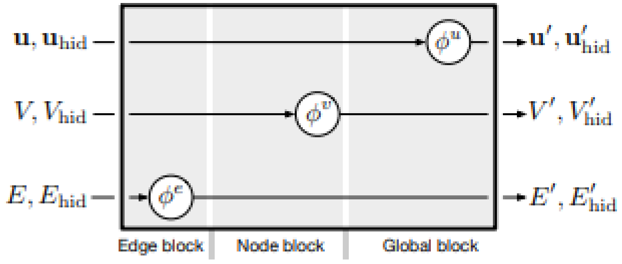
Compositionality: GN Blocks

GN block can have many different forms

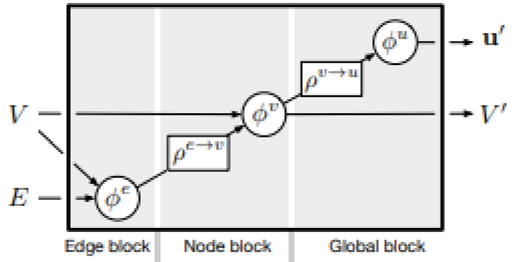
Recurrent



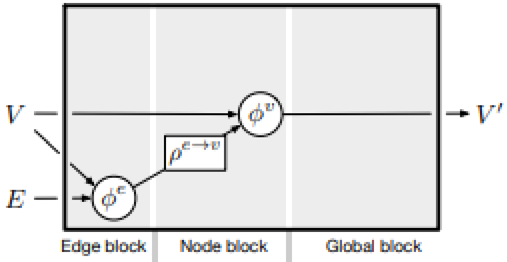
(a) Full GN block



(b) Independent recurrent block

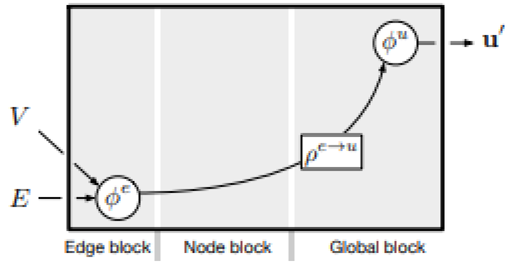


(c) Message-passing neural network

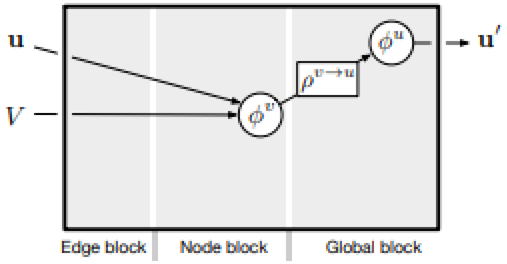


(d) Non-local neural network

Edges only



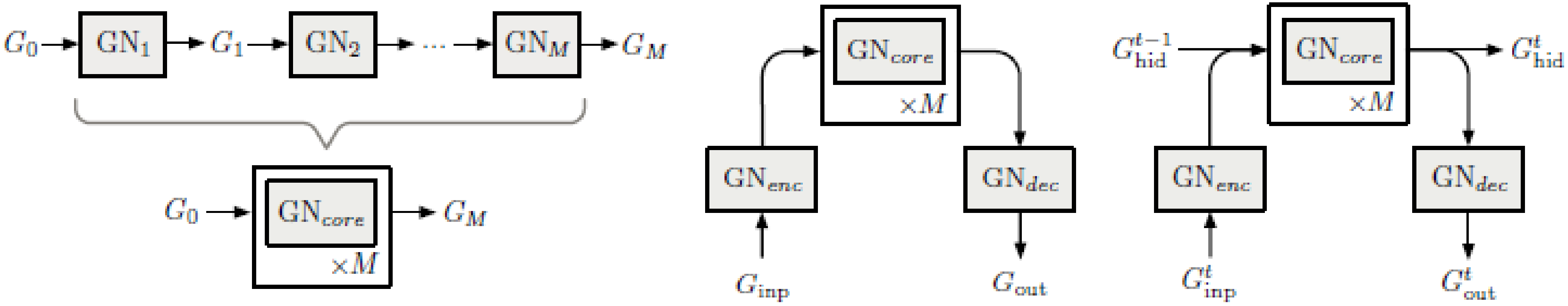
(e) Relation network



(f) Deep set

Compositionality: GN Blocks

If you stack GN blocks, you get compositionality (?)



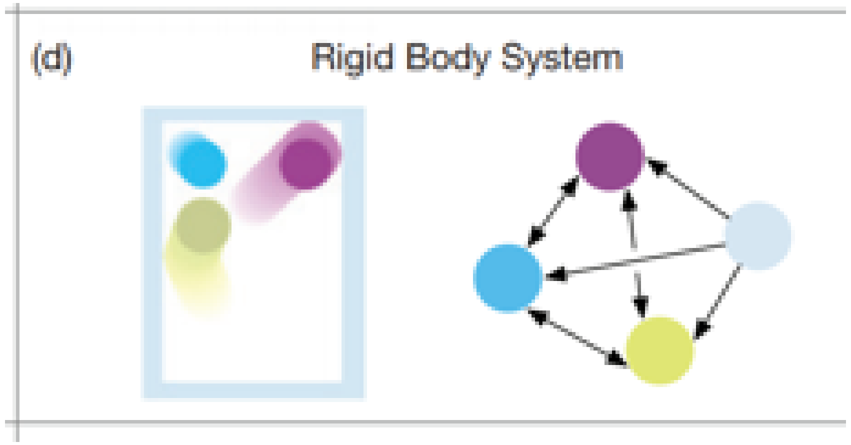
(a) Composition of GN blocks (b) Encode-process-decode (c) Recurrent GN architecture

Perhaps, being able to add/remove nodes was already a kind of compositionality?

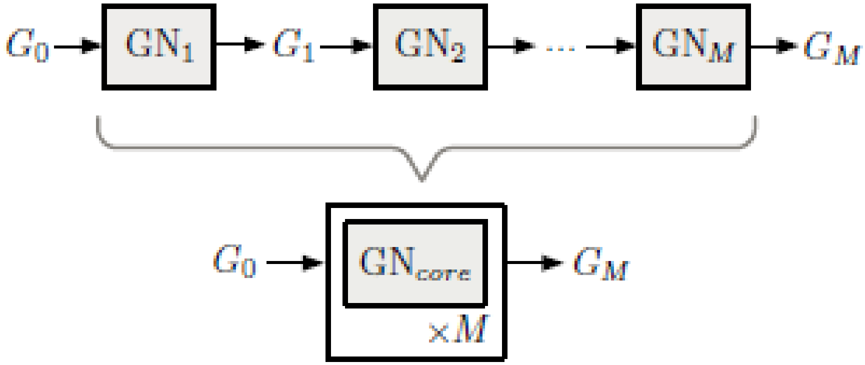
More blocks
Blocks with different properties
Change global physics, etc….

Compositionality: GN Blocks

I think their idea is that stack of GN blocks can learn progressively more abstract features?



Position? Velocity? Dynamics?



(a) Composition of GN blocks

Tensorflow Library for Graph Networks

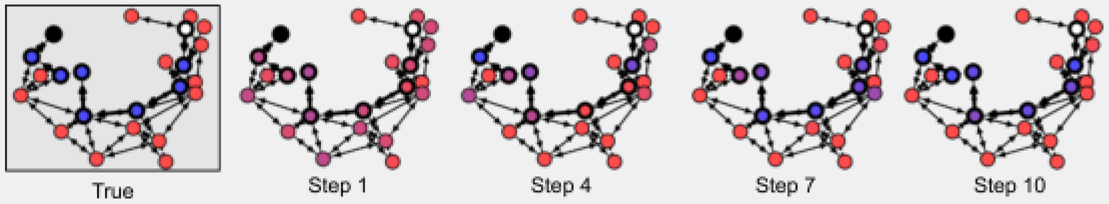
Box 4: Graph Nets open-source software library: `github.com/deepmind/graph_nets`

We have released an open-source library for building GNs in Tensorflow/Sonnet. It includes demos of how to create, manipulate, and train GNs to reason about graph-structured data, on a shortest path-finding task, a sorting task, and a physical prediction task. Each demo uses the same GN architecture, which highlights the flexibility of the approach.

Shortest path demo: `tinyurl.com/gn-shortest-path-demo`

This demo creates random graphs, and trains a GN to label the nodes and edges on the shortest path between any two nodes. Over a sequence of message-passing steps (as depicted by each step's plot), the model refines its prediction of the shortest path.

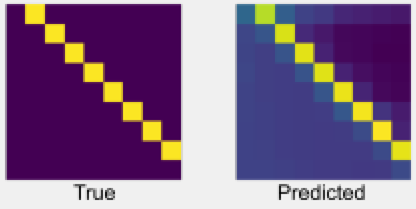
Shortest path: predictions at each message-passing step



Sort demo: `tinyurl.com/gn-sort-demo`

This demo creates lists of random numbers, and trains a GN to sort the list. After a sequence of message-passing steps, the model makes an accurate prediction of which elements (columns in the figure) come next after each other (rows).

Sort: item-to-item connections



Discussion

Discussion

1. How limited are the inductive biases in existing models? (e.g. CNN, RNN, L2, etc.).

2. If they're right, should everything be a graph NN, where we have studied special cases?

3. How would we use graph NNs in NLP? (say: compare to RNN). Parse trees?