Lecture 2: Graded modal logic and limits of GNN expressivity

Part 3: Graded modal logic and bisimulation

- How capabilities of GNNs be related to formal languages?
    - Formulas of graded modal logic define Boolean node properties of labelled graphs.
    - Compare properties of graphs that are definable by a formula vs computable by a GNN.
- What do we cover next?
    - Basics of graded modal logic.
    - Bisimulation, a central tool in modal logics, is essentially colour refinement.
    - We define characteristic formulae; these define a colour class of fixed round CR.
- Main sources:
    - Martin Otto: Graded modal logic and counting bisimulation. CoRR abs/1910.00039, 2019.

Why are we relating expressivity of GNNs to logics?

- Logic is a mature tool to study the expressivity of a model of computation.
- Automata theory: regular languages.
    - Languages accepted by finite automata and generated by regular expressions.
    - A first-order sentence interpreted over labelled linear orders

$$\mathrm{a} \to \mathrm{a} \to \mathrm{a} \to \mathrm{b} \to \mathrm{b}$$

$$\forall x \forall y \big((x \leq y \wedge B(x)) \to B(y)\big) \wedge \forall z (A(z) \vee B(z))$$

      defines the language $A^* B^*$.
    - First-order logic has the same uniform expressivity than star-free languages.
    - Monadic second-order logic has the same uniform expressivity than regular languages.
- Complexity theory: many complexity classes have logical counterparts.

Why are we relating expressivity of GNNs to logics?

- Logic is a mature tool to study the expressivity of a model of computation.
- Automata theory: regular languages.
- Complexity theory: many complexity classes have logical counterparts.
  - A property of ordered graphs can be decided in polynomial time if and only if it can be described in first-order logic + recursion.
  - A property of graphs can be decided in NP if and only if it can be described in existential second-order logic.

Why are we relating expressivity of GNNs to logics?

- Logic is a mature tool to study the expressivity of a model of computation.
- Automata theory: regular languages.
- Complexity theory: many complexity classes have logical counterparts.
  - A property of ordered graphs can be decided in polynomial time if and only if it can be described in first-order logic + recursion.
  - A property of graphs can be decided in NP if and only if it can be described in existential second-order logic.
- Can we understand the limits of uniform expressivity of GNNs using the logical lens?
- Field: Descriptive complexity of GNNs!

Modal logic is logic for labelled graphs

- We consider input graphs with abstract labels from some finite set.
- Fix a finite set of atomic propositions

$$P = \{p_1, \ldots, p_s\}.$$

- We consider graphs labeled by subsets of P:

$$g \colon N \to \mathcal{P}(P).$$

- Intuitively,

$$p \in g(n)$$

means that proposition p is true at node n.

Graded modal logic: syntax

The formulas of graded modal logic are generated by

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \psi \mid \varphi \vee \psi \mid \Diamond_{\geq k}\varphi,$$

where $p \in P$ and $k \geq 1$.

- The modality

  $$\Diamond_{\geq k}\varphi$$

  means: at least k neighbours satisfy φ.
- Ordinary modal logic is the special case where only $k = 1$ is used.
  We write $\Diamond\varphi$ for $\Diamond_{\geq 1}\varphi$.
- The modal depth of a formula is the maximum nesting depth of modalities.

Graded modal logic: semantics

Let $G = (N, E, g)$ be a $\mathcal{P}(P)$-labeled graph, with $\mathcal{P}(P)$ denoting the powerset of P.

- Atomic propositions are evaluated by labels:

$$G, n \models p \iff p \in g(n).$$

- Boolean connectives are interpreted as usual.
- The graded modality is interpreted by counting out-neighbours:

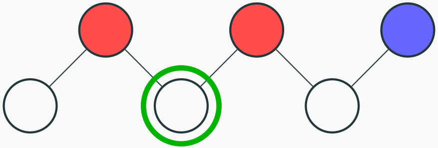
$$G, n \models \Diamond_{\geq k} \varphi$$

iff

$$\left| \{ m \in E(n) \mid G, m \models \varphi \} \right| \geq k.$$

GML formulas define node properties of labelled graphs

$$\varphi := \Diamond_{\geq 2}\mathrm{red}$$

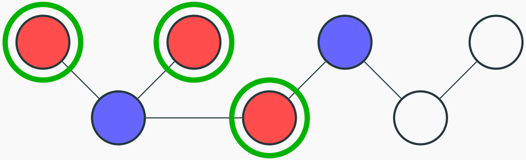


Green ring $= G, v \models \varphi$

"has at least two red neighbours"

GML formulas define node properties of labelled graphs

$$\psi := \Diamond_{\geq 1}\big(\text{purple} \wedge \Diamond_{\geq 2}\text{red}\big)$$



Green ring $= G, v \models \psi$

"has a purple neighbour that itself has at least two red neighbours"

GML formulas as node classifiers

Let φ be a GML formula.

- The formula φ defines a (boolean) node classifier

$$C_\varphi \colon \mathcal{G}[\mathcal{P}(P)] \to \mathcal{G}[\mathbb{B}].$$

- For every graph $G = (N, E, g)$, we set

$$C_\varphi(G) = (N, E, c_G^\varphi),$$

  where

$$c_G^\varphi \colon N \to \mathbb{B}.$$

- A node is classified as positive exactly when it satisfies φ:

$$c_G^\varphi(n) = 1 \quad \Longleftrightarrow \quad G, n \models \varphi.$$

From colour refinement to logic

- Colour refinement updates a node using

  its current colour   and   the multiset of colours of its neighbours.

- Graded modal logic describes the same kind of local counting information.
- A typical GML statement says:

  "there are at least k neighbours satisfying φ".

- Thus GML is a natural logical language for reasoning about colour refinement (CR)-types and message passing.

Logic and GNNs

- GML seems quite weak. How does it compare with GNNs?
- $\mathcal{P}(P)$-labeled graphs are transformed to $\mathbb{R}^k$ features using the one-hot encoding.
- What does your intuition say in regards to graphs with a finite set of labels?
  - Distinguishing power:
  - Uniform expressivity:
  - Non-uniform expressivity:

Logic and GNNs

- GML seems quite weak. How does it compare with GNNs?
- $\mathcal{P}(P)$-labeled graphs are transformed to $\mathbb{R}^k$ features using the one-hot encoding.
- What does your intuition say in regards to graphs with a finite set of labels?
  - Distinguishing power: GML $=$ GNN, we will prove this after the break!
  - Uniform expressivity:
  - Non-uniform expressivity:

Logic and GNNs

- GML seems quite weak. How does it compare with GNNs?
- $\mathcal{P}(P)$-labeled graphs are transformed to $\mathbb{R}^k$ features using the one-hot encoding.
- What does your intuition say in regards to graphs with a finite set of labels?
  - Distinguishing power: GML $=$ GNN, we will prove this after the break!
  - Uniform expressivity: GNNs are more expressive.
  - Non-uniform expressivity:

Logic and GNNs

- GML seems quite weak. How does it compare with GNNs?
- $\mathcal{P}(P)$-labeled graphs are transformed to $\mathbb{R}^k$ features using the one-hot encoding.
- What does your intuition say in regards to graphs with a finite set of labels?
  - Distinguishing power: GML $=$ GNN, we will prove this after the break!
  - Uniform expressivity: GNNs are more expressive. E.g., having more red neighbours than blue neighbours cannot be expressed in GML.
  - Non-uniform expressivity:

Logic and GNNs

- GML seems quite weak. How does it compare with GNNs?
- $\mathcal{P}(P)$-labeled graphs are transformed to $\mathbb{R}^k$ features using the one-hot encoding.
- What does your intuition say in regards to graphs with a finite set of labels?
  - Distinguishing power: GML = GNN, we will prove this after the break!
  - Uniform expressivity: GNNs are more expressive. E.g., having more red neighbours than blue neighbours cannot be expressed in GML.
  - Non-uniform expressivity: GML = GNN, when graph size is fixed. We will prove this after the break.

From formulas to equivalence

- GML formulas describe what can be observed from a node by looking locally along edges.
- This induces a natural question regarding invariance:

When do two pointed graphs satisfy the same GML formulas?

- The answer is given by graded bisimulation.
- Intuitively, two nodes are graded bisimilar if they have the same label and the same number of neighbours of each relevant type.
- Graded bisimulation is an equivalence relation. Hence, we can talk about properties of (pointed) labelled graphs that are g-bisimulation invariant.
- This is the logical counterpart of colour refinement.

Why bisimulation?

- Bisimulation compares pointed graphs by matching their local behaviour.
- For ordinary modal logic, it is enough to match successors back and forth.
- For graded modal logic, this is not enough: we must also match how many successors of each kind there are.
- Therefore we use graded bisimulation, also called counting bisimulation.

Idea of graded bisimulation

- A procedure to check whether two nodes are in the same colour class of CR.

(u, v) are graded bisimilar $u \sim v$, if their labels are identical, and

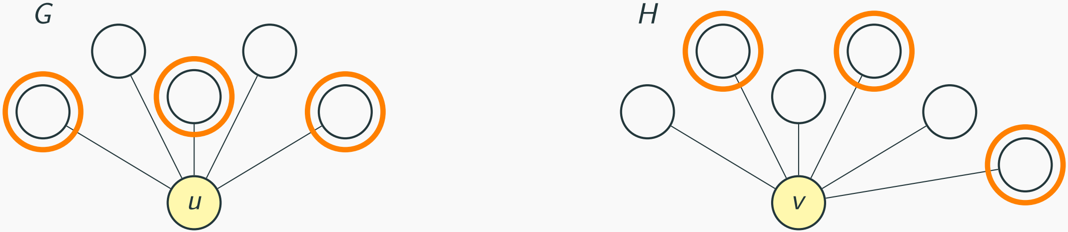
Idea of graded bisimulation

- A procedure to check whether two nodes are in the same colour class of CR.

for any sequence sequence (u_1, u_3, u_5) or successors of u

Idea of graded bisimulation

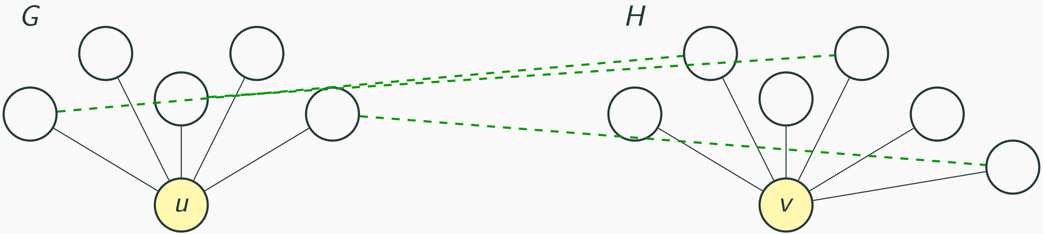
- A procedure to check whether two nodes are in the same colour class of CR.



there exists a sequence (v_2, v_4, v_6) of successors of v such that

Idea of graded bisimulation

- A procedure to check whether two nodes are in the same colour class of CR.



the corresponding elements are bisimilar

$u_1 \sim v_2, u_3 \sim v_4, u_5 \sim v_6$

Idea of graded bisimulation

- A procedure to check whether two nodes are in the same colour class of CR.

Similarly, for any sequence or successors of v there must be a corresponding sequence or successors of u

r-round graded bisimulation

Let (G, n) and (H, m) be pointed $\mathcal{P}(P)$-labeled graphs.

- We define equivalence relations $\sim_r$ recursively.
- At depth 0,
$$(G, n) \sim_0 (H, m) \iff g_G(n) = g_H(m).$$
  - Two nodes are equivalent at depth 0 iff they satisfy the same atomic propositions.
- At depth $r + 1$, we require the same label and the same number of successors of each r-round equivalence class.

r-round graded bisimulation: recursive step

We say

$$(G, n) \sim_{r+1} (H, m)$$

iff

$$g_G(n) = g_H(m)$$

and, for every $\sim_r$-class C,

$$|\{u \in E_G(n) \mid (G, u) \in C\}| = |\{v \in E_H(m) \mid (H, v) \in C\}|.$$

- In words: n and m have the same number of neighbours of every previous type.
- The equivalence class of (G, n) under $\sim_r$ is its r-round graded bisimulation type.

Intuition

- Depth 0 only sees the node label.
- Depth 1 sees the node label and counts the labels of its neighbours.
- Depth 2 sees the node label and counts the depth-1 types of its neighbours.
- In general, depth r sees the rooted neighbourhood up to depth r, but only through counting types.
- This is exactly the kind of information used by graded modal logic.

GML is invariant under graded bisimulation

Proposition 2

Let φ be a GML formula of modal depth at most r. If

$$(G, n) \sim_r (H, m),$$

then

$$G, n \models \varphi \iff H, m \models \varphi.$$

- Thus GML formulas of depth r cannot distinguish nodes with the same r-round graded bisimulation type.
- The proof is by induction on the structure of φ.

Connection to colour refinement

Proposition 3

Let χ_r be the colouring after r rounds of colour refinement. Then, for pointed graphs (G, n) and (H, m),

$$(G, n) \sim_r (H, m) \iff \chi_r^G(n) = \chi_r^H(m).$$

- Colour refinement computes graded bisimulation types.
- Two nodes have the same CR colour after r rounds exactly when they have the same r-round graded bisimulation type.
- The proof is by induction on r.

How expressive GML is?

- GML is expressive enough to define the r-round CR class of every pointed graph.
  - For every (G, n) there is a formula ψ s.t. $(H, m) \models \psi$ if and only if $\chi_r^G(n) = \chi_r^H(m)$.
  - We will prove this shortly.
- It does not follow that GML has the same uniform expressivity as GNNs. Why?

How expressive GML is?

- GML is expressive enough to define the r-round CR class of every pointed graph.
  - For every (G, n) there is a formula ψ s.t. $(H, m) \models \psi$ if and only if $\chi_r^G(n) = \chi_r^H(m)$.
  - We will prove this shortly.
- It does not follow that GML has the same uniform expressivity as GNNs. Why?
  - Let M_{GNN} be a node-feature map defined by an r-layer GNN.
  - Then, $\{(G, n) \mid M_{\mathsf{GNN}}(G)(n) = 1\}$ is invariant under r-round CR.
  - $\{(G, n) \mid M_{\mathsf{GNN}}(G)(n) = 1\}$ is a union of r-round C equivalence classes.
  - Each class is definable with a GML formula,

How expressive GML is?

- GML is expressive enough to define the r-round CR class of every pointed graph.
  - For every (G, n) there is a formula ψ s.t. $(H, m) \models \psi$ if and only if $\chi_r^G(n) = \chi_r^H(m)$.
  - We will prove this shortly.
- It does not follow that GML has the same uniform expressivity as GNNs. Why?
  - Let M_{GNN} be a node-feature map defined by an r-layer GNN.
  - Then, $\{(G, n) \mid M_{\mathsf{GNN}}(G)(n) = 1\}$ is invariant under r-round CR.
  - $\{(G, n) \mid M_{\mathsf{GNN}}(G)(n) = 1\}$ is a union of r-round C equivalence classes.
  - Each class is definable with a GML formula, but there can be infinitely many classes!

Characteristic formulae

- Fix a modal depth r.
- The r-type of a pointed graph (G, n) is its equivalence class under r-round graded bisimulation.
- A characteristic formula for this type is a GML formula

$$\vartheta^r_{G,n}$$

such that, for every pointed graph (H, m),

$$H, m \models \vartheta^r_{G,n} \quad \iff \quad (H, m) \sim_r (G, n).$$

- Thus $\vartheta^r_{G,n}$ describes exactly what can be seen from n up to depth r.

Characteristic formulae: depth 0

Let $G = (N, E, g)$ be a $\mathcal{P}(P)$-labeled graph.

- At depth 0, only the node label matters.
- For $n \in N$, define
$$\vartheta^0_{G,n} := \bigwedge_{p \in g(n)} p \ \wedge \bigwedge_{p \in P \setminus g(n)} \neg p.$$
- Then
$$H, m \models \vartheta^0_{G,n} \quad \Longleftrightarrow \quad g_H(m) = g_G(n).$$

Characteristic formulae: induction step

Suppose $\vartheta^r_{G,u}$ has been defined for all neighbours $u \in E(n)$.

- Let $\tau_1, \ldots, \tau_s$ be the distinct r-types occurring among the neighbours of n.
- Let k_i be the number of neighbours of n of type τ_i.
- Choose a characteristic formula $\vartheta^r_{\tau_i}$ for each τ_i.
- Then $\vartheta^{r+1}_{G,n}$ says:
    - the current node has the same labels as n;
    - it has exactly k_i neighbours of type τ_i;
    - every neighbour has one of these types.

Characteristic formulae: recursive definition

We may define

$$\vartheta^{r+1}_{G,n}$$

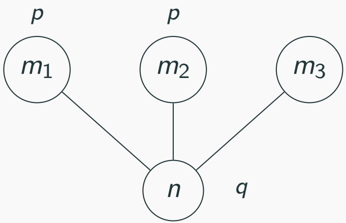
as

$$\vartheta^{0}_{G,n} \wedge \bigwedge_{i=1}^{s} \left(\Diamond_{\geq k_i} \vartheta^{r}_{\tau_i} \wedge \neg \Diamond_{\geq k_i+1} \vartheta^{r}_{\tau_i}\right)$$

$$\wedge \neg \Diamond \left(\bigwedge_{i=1}^{s} \neg \vartheta^{r}_{\tau_i} \right).$$

- The first conjunct fixes the node labels.
- The second part fixes the exact number of neighbours of each previous type.
- The last part says that there are no neighbours of any other type.

Example of a characteristic formula



- Take $r = 1$. Node n satisfies $q, \neg p$; among its neighbors, exactly two satisfy $p, \neg q$, and exactly one satisfies $\neg p, \neg q$.
- The 1-type of (G, n) is pinned down by counting, for each atomic type realized in $E(n)$, how many neighbors realize it.
- Characteristic formula:

$$\begin{aligned} \vartheta^1_{G,n} &= (q \wedge \neg p) \\ &\qquad \wedge \Diamond^{=2}(p \wedge \neg q) \wedge \Diamond^{=1}(\neg p \wedge \neg q) \\ &\qquad \wedge \neg\Diamond(\neg(p \wedge \neg q) \wedge \neg(\neg p \wedge \neg q)). \end{aligned}$$

- Exact counting can be writting with a $\geq k \ \wedge \ \neg{\geq}k{+}1$ pair.
- Any $(H, m) \models \vartheta^1_{G,n}$ is 1-bisimilar to (G, n): same label at m, same neighbor counts for combinations of p and q.

GML characterises local bisimulation types

Proposition 4

For every $r \in \mathbb{N}$ and every pointed graph (G, n), the characteristic formula

$$\vartheta^r_{G,n}$$

is a GML-formula of modal depth r such that

$$H, m \models \vartheta^r_{G,n} \iff (H, m) \sim_r (G, n).$$

- The proof is by induction on r, which we essentially already covered.
- Hence GML can define every local graded bisimulation type of a pointed graph.
- Since r-round graded bisimulation coincides with r-round CR equivalence, GML can also define every r-round CR-type of pointed graphs.

Toolkit: GML, bisimulation, and colour refinement

- Graded modal logic is a language for CR-type message passing.
- Bisimulation is an invariance from logic that corresponds to colour refinement.
- Charcteristic formulae can define all l-round CR-types.
- Next we show how these concepts can be used to yield results about GNNs.

Part 4: Logical characterisations of GNNs

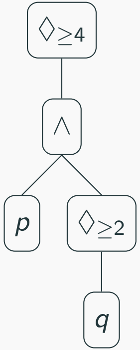
- Next we use tools from the previous lecture to study GNNs.
  - Distinguishability.
    When can two graphs be separated by a GNN?
  - Uniform expressivity:
    What are the properties that a single GNN can express over all graphs?
  - Non-uniform expressivity:
    What properties can be expressed if the used GNN can depend of size of the graph?
- What do we cover next?
  - We prove that GNNs are at least as expressive than GML.
    By showing how formulae of modal logic can be evaluated using a GNN.
  - We show that graded modal logic and GNNs have the same distinguishing power.
  - We prove that non-uniform expressivity of GNNs and GML coincide.
    Restricted to graphs of fixed size, GML can express all l-round CR colour classes.
- Main sources:
  - Pablo Barceló, Egor V. Kostylev, Mikaël Monet, Jorge Pérez, Juan L. Reutter, Juan Pablo Silva: The Logical Expressiveness of Graph Neural Networks. ICLR 2020.

Evaluating GML with GNNs

- A GML formula can be evaluated by a GNN by induction on its syntax tree, exemplified for $\Diamond_{\geq 4}(p \wedge \Diamond_{\geq 2} q)$.
- The feature vector of a node can store truth values of relevant subformulae:
  $(\Diamond_{\geq 4}(p \wedge \Diamond_{\geq 2} q), (p \wedge \Diamond_{\geq 2} q), \Diamond_{\geq 2} q, p, q)$.
- Atomic propositions are read from the input label.
- Boolean connectives are computed by feedforward neural networks.
- The graded modality

  $$\Diamond_{\geq k} \varphi$$

  is evaluated by summing the truth values of φ over the neighbours and checking whether the sum is at least k.



Syntax tree

Evaluating graded modalities

Suppose the current feature of a node records whether φ is true.

- For example, φ expresses "red" node p_{red}.
- Assume $G = (N, E, g)$ with $g : N \to \{0,1\}^3$ hot-one encoding of colours red, green and blue (red: $(1,0,0)$, green: $(0,1,0)$, blue: $(0,0,1)$).
- Aggregate the truth values over the neighbours:

$$s(n) := \sum_{m \in E(n)} (1 \cdot g(m)_1 + 0 \cdot g(m)_1 + 0 \cdot g(m)_3) = \sum_{m \in E(n)} g(m)_1$$

- Then we can express "having no less than k red nodes":

$$G, n \models \Diamond_{\geq k}\varphi \quad \Longleftrightarrow \quad \mathrm{trReLU}(s(n) - k - 1) = 1.$$

- A simple feedforward network can implement such threshold tests.
- Thus one message-passing layer can evaluate one modal step.

Simple homogeneous GNNs

- In general, different layers of a GNN may use different Agg and Comb functions
  - In homogeneous GNNs each layer is identical.

Simple homogeneous GNNs

- In general, different layers of a GNN may use different Agg and Comb functions
  - In homogeneous GNNs each layer is identical.
- In general, Agg and Comb function can be any functions.
  - Simple GNNs use (weighted) sums passed through non-linear activation function as Comb and Agg.

$$\mathrm{Agg}(X) := \sum_{\mathbf{x} \in \mathbf{X}} x$$

$$\mathrm{Comb}(\mathbf{x}, \mathbf{y}) := \sigma(\mathbf{x}\mathbf{C} + \mathbf{y}\mathbf{B} + \mathbf{b}),$$

where $A, C \in \mathbb{R}^{l \times l}$, $b \in \mathbb{R}^{l}$, and σ is the truncated ReLU activation defined by

$$\sigma(x) := \min(\max(0, x), 1).$$

Simple homogeneous GNNs

- In general, different layers of a GNN may use different Agg and Comb functions
  - In homogeneous GNNs each layer is identical.
- In general, Agg and Comb function can be any functions.
  - Simple GNNs use (weighted) sums passed through non-linear activation function as Comb and Agg.

$$\mathrm{Agg}(X) := \sum_{\mathbf{x} \in \mathbf{X}} x$$

$$\mathrm{Comb}(\mathbf{x}, \mathbf{y}) := \sigma(\mathbf{xC} + \mathbf{yB} + \mathbf{b}),$$

where $A, C \in \mathbb{R}^{l \times l}$, $b \in \mathbb{R}^{l}$, and σ is the truncated ReLU activation defined by

$$\sigma(x) := \min(\max(0, x), 1).$$

- For any GML formula, there is a particular choice of B, C, and b that are used to simulate the logical formula.

From GML formulae to GNNs

Proposition 5

For every GML formula φ, there is a fixed-depth simple homogeneous message-passing GNN that defines the same node classifier:

$$c_G^{\varphi}(n) = 1 \iff G, n \models \varphi.$$

- The depth of the GNN is controlled by the modal depth of φ.
  - Direct implementation yields depth of the syntax tree, which can be optimised to modal depth.
- The construction is by structural induction on φ.
- This gives a logical way to build GNNs from GML formulae.

Back to CR and GNNs: Proof via logic

Proposition 1 (Colour refinement and GNNs [Mor+19])

Let (G, n) and (H, m) be two pointed X-labeled graphs. Let χ_t denote the colouring after t rounds of colour refinement.

1. For every t-layer message-passing GNN with node features g_t, we have

$$\chi_t^G(n) = \chi_t^H(m) \implies g_t^G(n) = g_t^H(m).$$

2. Conversely, if colour refinement distinguishes the two pointed graphs after t rounds, that is,

$$\chi_t^G(n) \neq \chi_t^H(m),$$

then there exists a t-layer message-passing GNN such that

$$g_t^G(n) \neq g_t^H(m).$$

Proof of Proposition 2: harder direction

- No need to prove directly that GNNs can simulate CR, we did the required work already!

Proof.

Proof of Proposition 2: harder direction

- No need to prove directly that GNNs can simulate CR, we did the required work already!

Proof.

- Let (G, n) and (H, m) be such that $\chi_t^G(n) \neq \chi_t^H(m)$.
- By Propositions 3 and 4 the characteristic formula $\vartheta_{G,n}^t$ of modal depth t separates the pointed graphs

$$(G, n) \models \vartheta_{G,n}^t \quad \text{and} \quad (H, m) \not\models \vartheta_{G,n}^t.$$

Proof of Proposition 2: harder direction

- No need to prove directly that GNNs can simulate CR, we did the required work already!

Proof.

- Let (G, n) and (H, m) be such that $\chi_t^G(n) \neq \chi_t^H(m)$.
- By Propositions 3 and 4 the characteristic formula $\vartheta_{G,n}^t$ of modal depth t separates the pointed graphs

$$(G, n) \models \vartheta_{G,n}^t \quad \text{and} \quad (H, m) \not\models \vartheta_{G,n}^t.$$

- We can then find a simple homogeneous GNN with t layers that separates (G, n) and (H, m), by Proposition 5.

$\square$

Logical equivalence and distinguishability

We are ready to show that the distinguising powers of GML and GNNs coincide.

Proposition 6 (GNNs and logical equivalence)

Let (G, n) and (H, m) be two pointed X-labeled graphs. TFAE

1. *There exists t-layer message-passing GNN that separates (G, n) and (H, m).*
2. $(H, m) \models \neg\vartheta^t_{G,n}$.
3. *There exists a GML-formula φ of modal depth at most t s.t $(G, n) \models \varphi$ and $(H, m) \not\models \varphi$.*

Logical equivalence and distinguishability

We are ready to show that the distinguishing powers of GML and GNNs coincide.

Proposition 6 (GNNs and logical equivalence)

Let (G, n) and (H, m) be two pointed X-labeled graphs. TFAE

1. *There exists t-layer message-passing GNN that separates (G, n) and (H, m).*
2. $(H, m) \models \neg\vartheta^t_{G,n}$.
3. *There exists a GML-formula φ of modal depth at most t s.t $(G, n) \models \varphi$ and $(H, m) \not\models \varphi$.*

Proof.

(1 $\Leftrightarrow$ 2) By Proposition 1, 1. is equivalent to colour refinement yielding $\chi^G_t(n) \neq \chi^H_t(m)$, which is equivalent to $(H, m) \models \neg\vartheta^r_{G,n}$ by Propositions 3 and 4.
(2 $\Rightarrow$ 3) Trivial. (3 $\Rightarrow$ 2) Follows from Propositions 2 and 4. $\square$

A more detailed proof

$(1 \Rightarrow 2)$ Assume that there is a node-feature map M computed by a t-layer GNN that separates (G, m) and (H, m).

A more detailed proof

$(1 \Rightarrow 2)$ Assume that there is a node-feature map M computed by a t-layer GNN that separates (G, m) and (H, m). By Prop. 1, M is invariant under t-layer CR and hence $\chi_t^G(m) \neq \chi_t^H(m)$.

A more detailed proof

$(1 \Rightarrow 2)$ Assume that there is a node-feature map M computed by a t-layer GNN that separates (G, m) and (H, m). By Prop. 1, M is invariant under t-layer CR and hence $\chi_t^G(m) \neq \chi_t^H(m)$. By Prop. 3, $(G, n) \not\sim_t (H, m)$ and hence by Prop 4. $(H, m) \not\models \vartheta^t_{G,n}$.

A more detailed proof

$(1 \Rightarrow 2)$ Assume that there is a node-feature map M computed by a t-layer GNN that separates (G, m) and (H, m). By Prop. 1, M is invariant under t-layer CR and hence $\chi_t^G(m) \neq \chi_t^H(m)$. By Prop. 3, $(G, n) \not\sim_t (H, m)$ and hence by Prop 4. $(H, m) \not\models \vartheta^t_{G,n}$.

$(2 \Rightarrow 1)$ All the steps of the proof go both directions.

A more detailed proof

($1 \Rightarrow 2$) Assume that there is a node-feature map M computed by a t-layer GNN that separates (G, m) and (H, m). By Prop. 1, M is invariant under t-layer CR and hence $\chi_t^G(m) \neq \chi_t^H(m)$. By Prop. 3, $(G, n) \not\sim_t (H, m)$ and hence by Prop 4. $(H, m) \not\models \vartheta_{G,n}^t$.

($2 \Rightarrow 1$) All the steps of the proof go both directions.

($2 \Rightarrow 3$) Assume $(H, m) \models \neg\vartheta_{G,n}^t$ and notice that $(G, n) \models \vartheta_{G,n}^t$.

A more detailed proof

$(1 \Rightarrow 2)$ Assume that there is a node-feature map M computed by a t-layer GNN that separates (G, m) and (H, m). By Prop. 1, M is invariant under t-layer CR and hence $\chi_t^G(m) \neq \chi_t^H(m)$. By Prop. 3, $(G, n) \not\sim_t (H, m)$ and hence by Prop 4. $(H, m) \not\models \vartheta^t_{G,n}$.

$(2 \Rightarrow 1)$ All the steps of the proof go both directions.

$(2 \Rightarrow 3)$ Assume $(H, m) \models \neg\vartheta^t_{G,n}$ and notice that $(G, n) \models \vartheta^t_{G,n}$.

$(3 \Rightarrow 2)$ Let φ be a GML-formula of modal depth at most t s.t $(G, n) \models \varphi$ and $(H, m) \not\models \varphi$.

A more detailed proof

($1 \Rightarrow 2$) Assume that there is a node-feature map M computed by a t-layer GNN that separates (G, m) and (H, m). By Prop. 1, M is invariant under t-layer CR and hence $\chi_t^G(m) \neq \chi_t^H(m)$. By Prop. 3, $(G, n) \not\sim_t (H, m)$ and hence by Prop 4. $(H, m) \not\models \vartheta^t_{G,n}$.

($2 \Rightarrow 1$) All the steps of the proof go both directions.

($2 \Rightarrow 3$) Assume $(H, m) \models \neg\vartheta^t_{G,n}$ and notice that $(G, n) \models \vartheta^t_{G,n}$.

($3 \Rightarrow 2$) Let φ be a GML-formula of modal depth at most t s.t $(G, n) \models \varphi$ and $(H, m) \not\models \varphi$. By Proposition 2 it follows that $(G, n) \not\sim_t (H, m)$. Proposition 4 then yields $(H, m) \models \neg\vartheta^t_{G,n}$.

Implications to GNN-architecture selection

- GNNs have the same distinguishing power than GML.
- GML formulas can be simulated by simple homogeneous GNNs
- Simple homogeneous GNNs have the same distinguishing power as all GNNs!

GNNs and GML have the same non-uniform expressivity

Let's consider the non-uniform case, where GNNs may depend on the the graph size:

Proposition 7

For every sequence of fixed depth message-passing GNNs $(M_s)_{s\in\mathbb{N}}$ that compute Boolean node classifiers there exists a sequence of GML formulae $(\varphi_s)_{s\in\mathbb{N}}$ such that for every labelled pointed graph (G, n)

$$M_s(G, n) = 1 \textit{ if and only if } G, n \models \varphi_s,$$

where $s = |G|$, and vice versa.

Proof.

Fix $s \in \mathbb{N}$ and t be the number of layers of M_s. By Props. 1 and 3, the model of M_s is invariant under t-round graded bisimulation. Set φ_s to be the disjunction of those $\psi^t_{G,n}$ s.t. $|G| = s$ and $M_s(G, n) = 1$. By Prop. 4, the disjuncts of φ_s characterise the bisimulation equivalence classes where $M_s(G, n) = 1$. Vice versa by Prop. 5. $\square$

Proof of Proposition 7 in more detail

- Let M_s be a t-layer GNN that is used to decide G, n of size s.
- We have shown that M_s is invariant under t-round CR.
- Restricted to graphs of size s, there are only finitely many t-round CR types. Easy to proof by induction on t.
- We have shown that each of those t-round CR types are infact, t-round graded bisimulation types.
- For each of those finite types, put a representative pair (G', m') to a set Φ, if $M_s(G', m') = 1$.
- Claim:
$$M_s(G, n) = 1 \text{ if and only if } G, n \models \bigvee_{(G', n') \in \Phi} \vartheta^t_{G', n'},$$
where $\vartheta^r_{G', m'}$ is the t-characteristic formula of G, n.
- Proposition 6 shows that vice versa holds even without restricting the graph size.

Uniform expressivity

- GNNs can simulate logical formulae
  ($\Rightarrow$) GNNs are at least as expressive than GML.
- GML is invariant under so-called bounded colour refinement
  - Consider *CR* that, for some fixed k, does not differentiate multiplicities larger than k
  - $\chi_{t+1}(n) = \chi_{t+1}(m)$, if $\chi_t(n) = \chi_t(m)$ and, for any colour, m and n have the same number of neighbours of that colour, or both have at least k neighbours of that colour.

Uniform expressivity

- GNNs can simulate logical formulae
  ($\Rightarrow$) GNNs are at least as expressive than GML.
- GML is invariant under so-called bounded colour refinement
  - Consider *CR* that, for some fixed k, does not differentiate multiplicities larger than k
  - $\chi_{t+1}(n) = \chi_{t+1}(m)$, if $\chi_t(n) = \chi_t(m)$ and, for any colour, m and n have the same number of neighbours of that colour, or both have at least k neighbours of that colour.
  - If node property is definable by a GML formula with modal depth d and graded modalities $\Diamond_{\geq i}$ where $i \leq k$, then the node property is invariant under d-round k-bounded CR.
- Node property: more neighbours satisfying p than neighbours satisfying q is not invariant under bounded CR, and thus not in GML.

Uniform expressivity

- GNNs can simulate logical formulae
  ($\Rightarrow$) GNNs are at least as expressive than GML.
- GML is invariant under so-called bounded colour refinement
  - Consider *CR* that, for some fixed k, does not differentiate multiplicities larger than k
  - $\chi_{t+1}(n) = \chi_{t+1}(m)$, if $\chi_t(n) = \chi_t(m)$ and, for any colour, m and n have the same number of neighbours of that colour, or both have at least k neighbours of that colour.
  - If node property is definable by a GML formula with modal depth d and graded modalities $\Diamond_{\geq i}$ where $i \leq k$, then the node property is invariant under d-round k-bounded CR.
- Node property: more neighbours satisfying p than neighbours satisfying q is not invariant under bounded CR, and thus not in GML.
- Easy to design a GNN for this property.
- GNNs are at strictly more expressive than GML.

Overview of the non-recursive case

- GNNs and GML have the same distinguishing power. This is captured by CR, or equivalently by graded bisimulation.
- When parameterised with the size of the input graph, GNNs and GML have the same non-uniform expressivity.
- In the uniform case, GNNs are strictly more expressive than GML.
- Literature gives some cases where the uniform expressivity of GNNs can be exactly characterised with a logic.
  - In these logics are enriched with some mild form of unbounded counting such as counting terms t and comparisons between those terms $t \leq t'$:

    $$t ::= \Diamond_{=x}\varphi,$$

    returns the number of neighbours satisfying φ.
- There are simple properties that cannot be uniformly expressed with GNNs
  - Not-invariant under CR (triangle detection) or fixed round CR (reachability).