Introduction to Graph Learning

Bastian Rieck (@Pseudomanifold)

What is a graph?

A graph is a tuple (V, E), consisting of a set of vertices V and a set of edges E, consisting of subsets of paired vertices.

What is a graph?

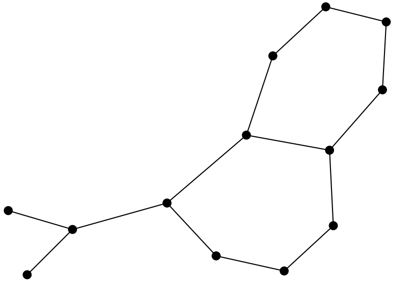
A graph is a tuple (V, E), consisting of a set of vertices V and a set of edges E, consisting of subsets of paired vertices.

If E is *ordered* instead, the graph is called *directed*.

What is a graph?

A graph is a tuple (V, E), consisting of a set of vertices V and a set of edges E, consisting of subsets of paired vertices.

If E is *ordered* instead, the graph is called *directed*.



Alternative views on graphs

A graph is a *triangulation* of a manifold.

Alternative views on graphs

A graph is a *triangulation* of a manifold.

A graph is a 1-dimensional simplicial complex.

Alternative views on graphs

A graph is a *triangulation* of a manifold.

A graph is a 1-dimensional simplicial complex.

A graph is a metric space.

Alternative views on graphs

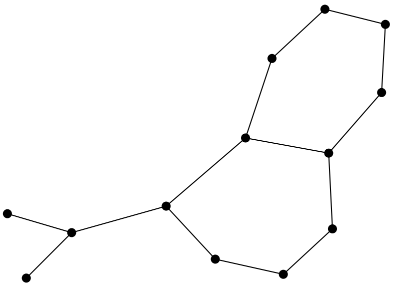
A graph is a *triangulation* of a manifold.

A graph is a 1-dimensional simplicial complex.

A graph is a metric space.

A graph is a set system.

What is graph learning?



Tasks

- Graph/node/edge classification
- Graph/node/edge regression
- Link prediction

Where do graphs come from?

A collection of different attitudes

Alignment	Belief

Where do graphs come from?

A collection of different attitudes

Alignment	Belief
Lawful	Graphs occur only in graph theory.

Where do graphs come from?

A collection of different attitudes

Alignment	Belief
Lawful	Graphs occur only in graph theory.
Neutral	Graphs can arise from other data modalities.

Where do graphs come from?

A collection of different attitudes

Alignment	Belief
Lawful	Graphs occur only in graph theory.
Neutral	Graphs can arise from other data modalities.
Chaotic	Everything is a graph.

Where do graphs come from?

A collection of different attitudes

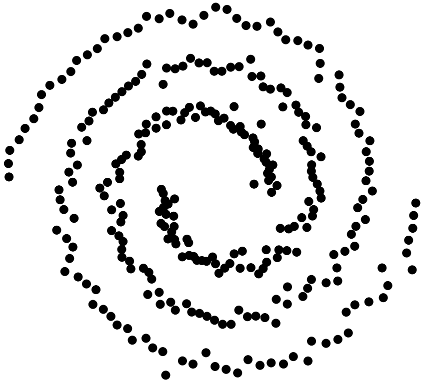
Alignment	Belief
Lawful	Graphs occur only in graph theory.
Neutral	Graphs can arise from other data modalities.
Chaotic	Everything is a graph.

Most graph theorists will agree that among the vast number of graphs that exist there are only a few thousand that can be considered really interesting.

(`https://houseofgraphs.org`)

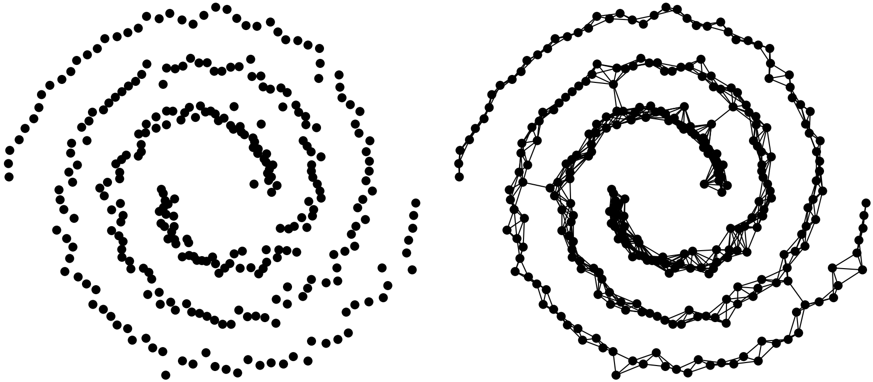
Obtaining graphs from other modalities

Point clouds



Obtaining graphs from other modalities

Point clouds

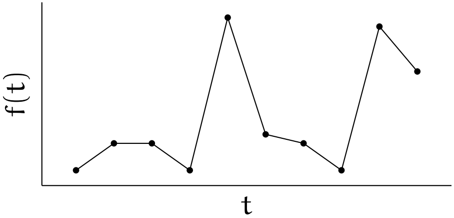


Rips graph at scale ϵ

$\mathcal{R}_\epsilon := (X, E)$ with $E := \{x, y \in X \mid d(x, y) \leqslant \epsilon\}$

Obtaining graphs from other modalities

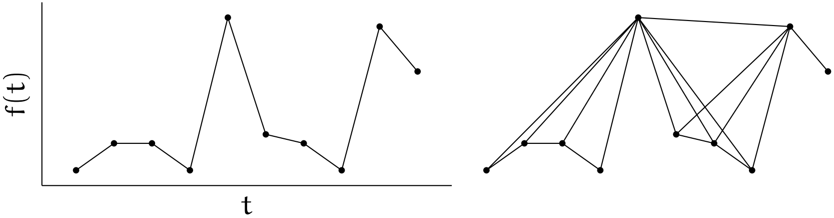
Time series



[1] L. Lacasa, B. Luque, F. Ballesteros, J. Luque and J. C. Nuño, 'From time series to complex networks: The visibility graph', *Proceedings of the National Academy of Sciences* 105.13, 2008, pp. 4972–4975.

Obtaining graphs from other modalities

Time series



Visibility graph[1]

Connect observations (t_i, f_i) and (t_{i+1}, f_{i+1}) if no other observations occur along their linear interpolation.

[1] L. Lacasa, B. Luque, F. Ballesteros, J. Luque and J. C. Nuño, 'From time series to complex networks: The visibility graph', *Proceedings of the National Academy of Sciences* 105.13, 2008, pp. 4972–4975.

How do these graphs *differ* from other graphs?

Some graphs arise from some prior *geometry*.

How do these graphs *differ* from other graphs?

Some graphs arise from some prior *geometry*.

Attributes in such graphs can reflect said geometry.

How do these graphs *differ* from other graphs?

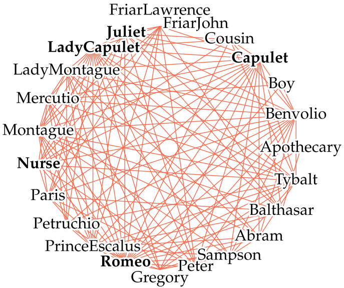
Some graphs arise from some prior *geometry*.

Attributes in such graphs can reflect said geometry.

When studying the graph, we are actually studying its geometry.

What's in a graph?

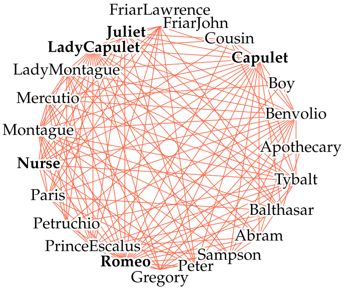
C. Coupette, J. Vreeken and **B. Rieck**, 'All the World's a (Hyper)Graph: A Data Drama', 2022, arXiv: `2206.08225 [cs.LG]`, URL: `https://hyperbard.net`

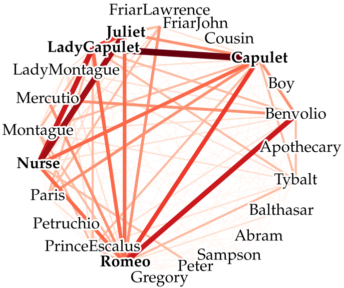


Three *valid* co-occurrence networks of characters in Shakespeare's Romeo and Juliet. Characters in Act III, Scene V are highlighted.

What's in a graph?

C. Coupette, J. Vreeken and **B. Rieck**, 'All the World's a (Hyper)Graph: A Data Drama', 2022, arXiv: `2206.08225 [cs.LG]`, URL: `https://hyperbard.net`

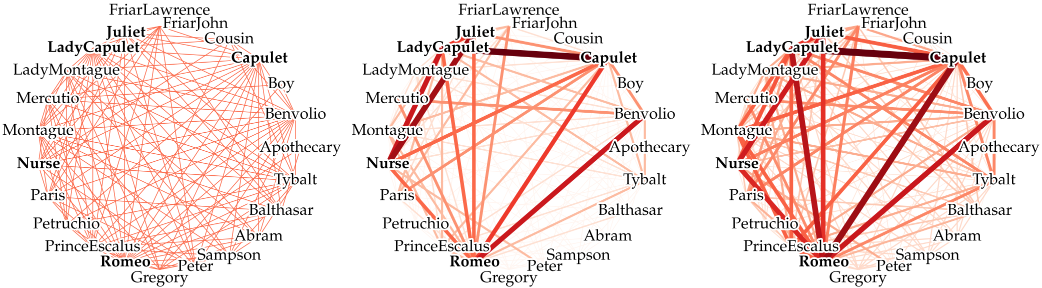




Three *valid* co-occurrence networks of characters in Shakespeare's Romeo and Juliet. Characters in Act III, Scene V are highlighted.

What's in a graph?

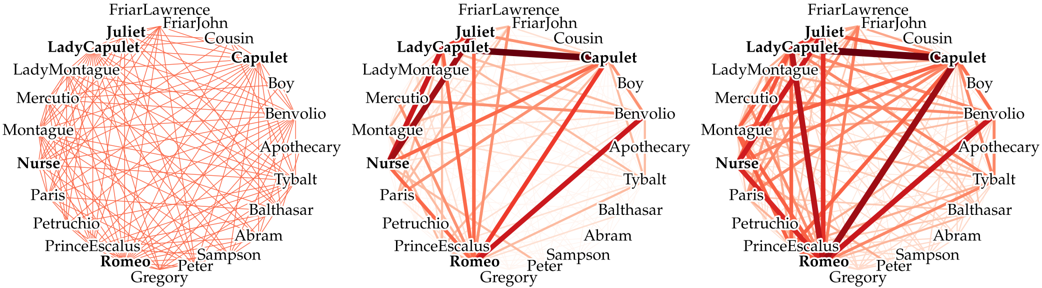
C. Coupette, J. Vreeken and **B. Rieck**, 'All the World's a (Hyper)Graph: A Data Drama', 2022, arXiv: `2206.08225 [cs.LG]`, URL: `https://hyperbard.net`



Three *valid* co-occurrence networks of characters in Shakespeare's Romeo and Juliet. Characters in Act III, Scene V are highlighted.

What's in a graph?

C. Coupette, J. Vreeken and **B. Rieck**, 'All the World's a (Hyper)Graph: A Data Drama', 2022, arXiv: `2206.08225 [cs.LG]`, URL: `https://hyperbard.net`



Three *valid* co-occurrence networks of characters in Shakespeare's Romeo and Juliet. Characters in Act III, Scene V are highlighted.

Observation

Romeo and the Capulets almost never interact directly; our *modelling decision* introduces new information!

How to represent graphs?

Two graphs G and G' can have a *different* number of vertices.

Hence, we require a *vectorised representation* $f\colon \mathcal{G} \to \mathbb{R}^d$ of graphs.

Such a representation f needs to be *permutation-invariant*.

Now and then

Shallow approaches

- node2vec (encoder–decoder)
- Graph kernels (RKHS feature maps)
- Laplacian-based embeddings

Deep approaches

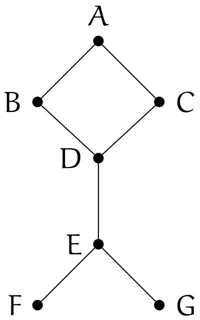
- Graph convolutional networks
- Graph isomorphism networks
- Graph attention networks

Message passing

The predominant paradigm in graph machine learning

Concept

Neighbouring nodes can exchange *messages* x, y, z (vectors in $\mathbb{R}^d$), which are *aggregated* (via a sum, a mean, or other permutation-invariant functions).

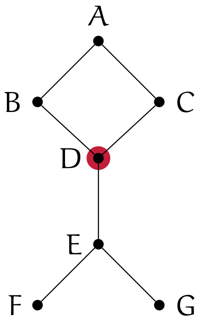


Message passing

The predominant paradigm in graph machine learning

Concept

Neighbouring nodes can exchange *messages* x, y, z (vectors in $\mathbb{R}^d$), which are *aggregated* (via a sum, a mean, or other permutation-invariant functions).

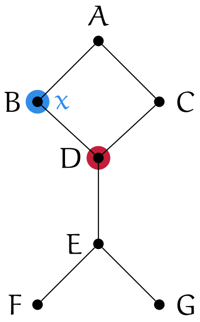


Message passing

The predominant paradigm in graph machine learning

Concept

Neighbouring nodes can exchange *messages* x, y, z (vectors in $\mathbb{R}^d$), which are *aggregated* (via a sum, a mean, or other permutation-invariant functions).

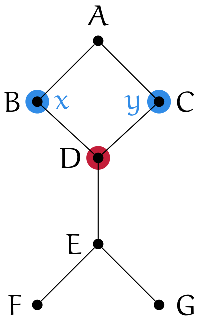


Message passing

The predominant paradigm in graph machine learning

Concept

Neighbouring nodes can exchange *messages* x, y, z (vectors in $\mathbb{R}^d$), which are *aggregated* (via a sum, a mean, or other permutation-invariant functions).

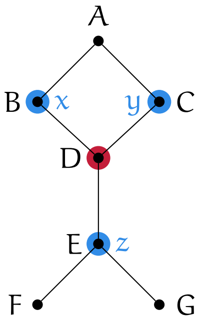


Message passing

The predominant paradigm in graph machine learning

Concept

Neighbouring nodes can exchange *messages* x, y, z (vectors in $\mathbb{R}^d$), which are *aggregated* (via a sum, a mean, or other permutation-invariant functions).

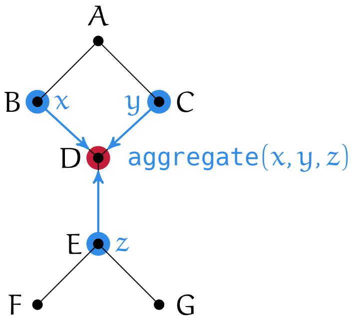


Message passing

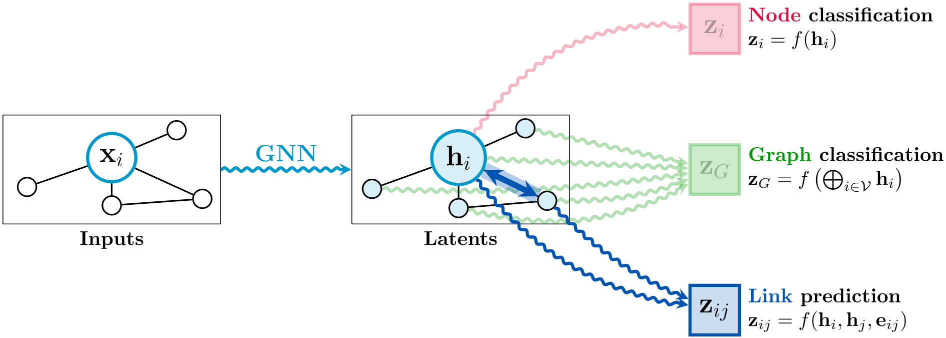
The predominant paradigm in graph machine learning

Concept

Neighbouring nodes can exchange *messages* x, y, z (vectors in $\mathbb{R}^d$), which are *aggregated* (via a sum, a mean, or other permutation-invariant functions).



Schematic overview of graph neural networks



Original source: P. Veličković, 'Everything is connected: Graph neural networks', *Current Opinion in Structural Biology* 79, 2023, p. 102538

Defining a graph neural network architecture in practice

Let ϕ, ψ *neural networks*, for example $\psi(\mathbf{x}) := \mathrm{ReLU}(\mathbf{Wx} + \mathbf{b})$, and $\bigoplus$ be *any* permutation-invariant aggregation function:

$$\text{Convolutional} \qquad \mathbf{h}_u = \phi\left(\mathbf{x}_u, \bigoplus_{v \in \mathcal{N}_u} c_{vu}\psi(\mathbf{x}_v)\right)$$

$$\text{Attentional} \qquad \mathbf{h}_u = \phi\left(\mathbf{x}_u, \bigoplus_{v \in \mathcal{N}_u} a(\mathbf{x}_u, \mathbf{x}_v)\psi(\mathbf{x}_v)\right)$$

$$\text{General} \qquad \mathbf{h}_u = \phi\left(\mathbf{x}_u, \bigoplus_{v \in \mathcal{N}_u} \psi(\mathbf{x}_u, \mathbf{x}_v)\right)$$

Expressive power increases from top to bottom! See M. M. Bronstein, J. Bruna, T. Cohen and P. Veličković, 'Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges', 2021, arXiv: `2104.13478 [cs.LG]` for more information on these architectures.

Towards measuring expressivity

Graph isomorphism

Given two graphs G, G' with vertices V, V', a *graph isomorphism* is a bijection $f\colon V \to V'$ such that u and v are adjacent in G *if and only if* $f(u)$ and $f(v)$ are adjacent in G'.

How to become famous: prove that the graph isomorphism problem can be solved in *polynomial time*.

Measuring expressivity

The Weisfeiler–Le(h)man test for graph isomorphism

1. *Create* a colour for each node in the graph (based on its label or its degree).

If the compressed labels of two graphs *diverge*, the graphs are *not* isomorphic!

The other direction is not valid! Non-isomorphic graphs can give rise to coinciding compressed labels.

Measuring expressivity

The Weisfeiler–Le(h)man test for graph isomorphism

1. *Create* a colour for each node in the graph (based on its label or its degree).
2. *Collect* colours of adjacent nodes in a multiset.

If the compressed labels of two graphs *diverge*, the graphs are *not* isomorphic!

The other direction is not valid! Non-isomorphic graphs can give rise to coinciding compressed labels.

Measuring expressivity

The Weisfeiler–Le(h)man test for graph isomorphism

1. *Create* a colour for each node in the graph (based on its label or its degree).
2. *Collect* colours of adjacent nodes in a multiset.
3. *Compress* the colours in the multiset and the node's colour to form a new one.

If the compressed labels of two graphs *diverge*, the graphs are *not* isomorphic!

The other direction is not valid! Non-isomorphic graphs can give rise to coinciding compressed labels.

Measuring expressivity

The Weisfeiler–Le(h)man test for graph isomorphism

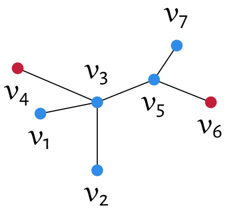
1. *Create* a colour for each node in the graph (based on its label or its degree).
2. *Collect* colours of adjacent nodes in a multiset.
3. *Compress* the colours in the multiset and the node's colour to form a new one.
4. *Continue* this relabelling scheme until the colours are stable.

If the compressed labels of two graphs *diverge*, the graphs are *not* isomorphic!

The other direction is not valid! Non-isomorphic graphs can give rise to coinciding compressed labels.

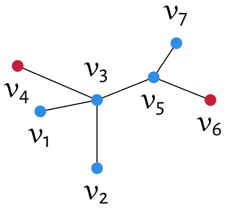
Weisfeiler–Le(h)man subtree features

Example for $h = 1$



Weisfeiler–Le(h)man subtree features

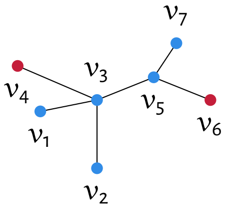
Example for $h = 1$



Node	Own label	Adjacent labels
v_1	🔵	🔵
v_2	🔵	🔵
v_3	🔵	🔵🔵🔵🔴
v_4	🔴	🔵
v_5	🔵	🔵🔵🔴
v_6	🔴	🔵
v_7	🔵	🔵

Weisfeiler–Le(h)man subtree features

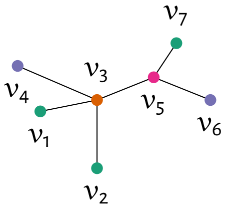
Example for $h = 1$



Node	Own label	Adjacent labels	Hashed label
v_1	🔵	🔵	🟢
v_2	🔵	🔵	🟢
v_3	🔵	🔵🔵🔵🔴	🟠
v_4	🔴	🔵	🟣
v_5	🔵	🔵🔵🔴	🩷
v_6	🔴	🔵	🟣
v_7	🔵	🔵	🟢

Weisfeiler–Le(h)man subtree features

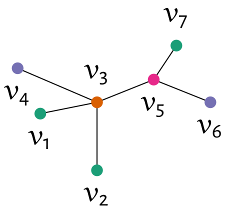
Example for $h = 1$



Node	Own label	Adjacent labels	Hashed label
v_1	●	●	●
v_2	●	●	●
v_3	●	● ● ● ●	●
v_4	●	●	●
v_5	●	● ● ●	●
v_6	●	●	●
v_7	●	●	●

Weisfeiler–Le(h)man subtree features

Example for $h = 1$



Label				
Count	3	1	2	1
Feature vector	$\Phi(G) := (3, 1, 2, 1)$			

Compare graphs based on *feature vectors*!

Expressivity

Surprising results

GNNs are no more expressive than the Weisfeiler–Le(h)man test for graph isomorphism.

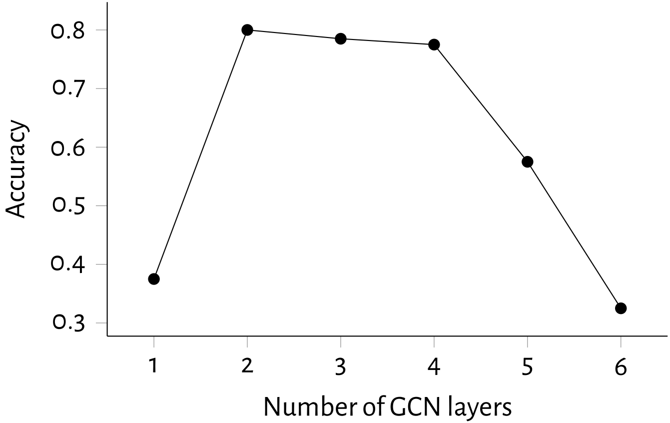
C. Morris et al., 'Weisfeiler and Leman Go Neural: Higher-Order Graph Neural Networks', *AAAI*, 2019

K. Xu, W. Hu, J. Leskovec and S. Jegelka, 'How Powerful are Graph Neural Networks?', *International Conference on Learning Representations (ICLR)*, 2019

Issues with graph neural networks

Oversmoothing

Having more GCN layers does not always result in higher performance. In fact, more layers can easily 'hide' the signal.



Issues with graph neural networks

Z. Chen, L. Chen, S. Villar and J. Bruna, ‘Can Graph Neural Networks Count Substructures?’, *Advances in Neural Information Processing Systems*, vol. 33, Curran Associates, Inc., 2020, pp. 10383–10395

Things GNNs *may* not be aware of

Number of connected components of a graph.

Issues with graph neural networks

Z. Chen, L. Chen, S. Villar and J. Bruna, 'Can Graph Neural Networks Count Substructures?', *Advances in Neural Information Processing Systems*, vol. 33, Curran Associates, Inc., 2020, pp. 10383–10395

Things GNNs *may* not be aware of

- Number of connected components of a graph.
- Number of certain substructures (triangles, 3-stars, ...) in a graph.

Issues with graph neural networks

Z. Chen, L. Chen, S. Villar and J. Bruna, 'Can Graph Neural Networks Count Substructures?', *Advances in Neural Information Processing Systems*, vol. 33, Curran Associates, Inc., 2020, pp. 10383–10395

Things GNNs *may* not be aware of

- Number of connected components of a graph.
- Number of certain substructures (triangles, 3-stars, …) in a graph.
- Number of cycles of arbitrary length in a graph.

The end

Graphs are often the right modality.

The end

Graphs are often the right modality.

The field is ever-growing!

The end

Graphs are often the right modality.

The field is ever-growing!

New frontiers: higher-order information, dynamics, expressivity, ...

The end

Graphs are often the right modality.

The field is ever-growing!

New frontiers: higher-order information, dynamics, expressivity, ...

Check out LoG, the 'Learning on Graphs' Conference. We also have reading groups!