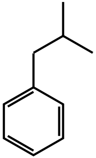
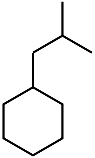
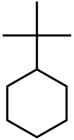
Measuring Change and Similarity of Graphs

Martin Grohe

Which of these graphs are most similar?

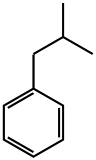




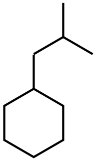


Which of these graphs are most similar ...

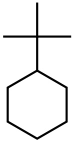
...when it comes to designing synthetic fuels?



Iso-Butylbenzene



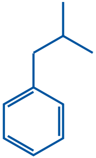
Iso-Butylcyclohexane



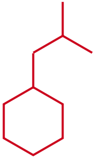
Tert-Butylcyclohexane

Which of these graphs are most similar . . .

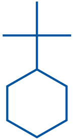
. . . when it comes to designing synthetic fuels?



Iso-Butylbenzene



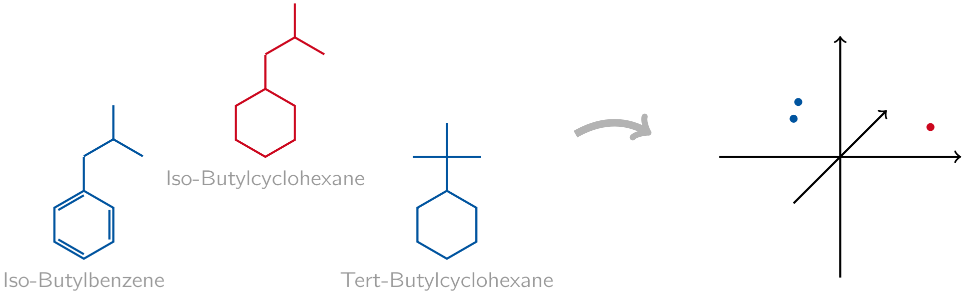
Iso-Butylcyclohexane



Tert-Butylcyclohexane

Which of these graphs are most similar ...

...when it comes to designing synthetic fuels?



Schweidtmann, Rittig, König, G., Mitsos, Dahmen, *Energy and Fuels 2021*

Machine Learning on Graphs

- machine learning is based on the promise that similar objects should have similar properties
- what "similar" means depends on the properties we are interested in

Why Should We Care?

Machine Learning on Graphs

- machine learning is based on the promise that similar objects should have similar properties
- what “similar” means depends on the properties we are interested in

Dynamical Processes on Graphs

- want to understand how a graph property evolves as the graph changes
- could be measured by difference quotients

$$\frac{\left|f(G) - f(\widetilde{G})\right|}{\delta\left(G, \widetilde{G}\right)}.$$

Why Should We Care?

Machine Learning on Graphs

- machine learning is based on the promise that similar objects should have similar properties
- what “similar” means depends on the properties we are interested in

Dynamical Processes on Graphs

- want to understand how a graph property evolves as the graph changes
- could be measured by difference quotients

$$\frac{\left|f(G)-f(\widetilde{G})\right|}{\delta\left(G,\widetilde{G}\right)}.$$

distance between G and $\widetilde{G}$

class of all graphs

A graph metric is an isomorphism-invariant mapping $\delta : \mathscr{G} \times \mathscr{G} \to \mathbb{R}_{\geq 0}$ such that for all graphs G, H, I:

(i) $G \cong H \implies \delta(G, H) = 0$;

(ii) $\delta(G, H) = \delta(H, G)$;

(iii) $\delta(G, I) \leq \delta(G, H) + \delta(H, I)$.

$$G \cong G', H \cong H' \implies \delta(G,H) = \delta(G',H')$$

A graph metric is an isomorphism-invariant mapping $\delta : \mathcal{G} \times \mathcal{G} \to \mathbb{R}_{\geq 0}$ such that for all graphs G, H, I:

(i) $G \cong H \implies \delta(G,H) = 0$;

(ii) $\delta(G,H) = \delta(H,G)$;

(iii) $\delta(G,I) \leq \delta(G,H) + \delta(H,I)$.

A graph metric is an isomorphism-invariant mapping $\delta : \mathscr{G} \times \mathscr{G} \to \mathbb{R}_{\geq 0}$ such that for all graphs G, H, I:

(i) $G \cong H \implies \delta(G, H) = 0$;

(ii) $\delta(G, H) = \delta(H, G)$;

(iii) $\delta(G, I) \leq \delta(G, H) + \delta(H, I)$.

It would be more precise to speak of a graph pseudo metric, and only of a metric if the converse of (i) holds as well (but we don't).

Measuring Similarity

We give no formal definition of what constitutes a "similarity measure".

We give no formal definition of what constitutes a “similarity measure”.

Similarity measures can be derived from metrics by applying an anti-monotone function, for example,

$$\sigma(G,H) := \exp\bigl(-c\cdot\delta(G,H)\bigr)$$

for a graph metric δ and a constant $c>0$.

We start by defining several metrics on graphs G, H of the same order $|G| = |H| = n$.
W.l.o.g. $V(G) = V(H) = [n]$.

We start by defining several metrics on graphs G, H of the same order $|G| = |H| = n$. W.l.o.g. $V(G) = V(H) = [n]$.

Graph Edit Distance

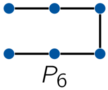
$$\delta_{\mathrm{ed}}(G, H) := \min\left\{ |D| \;\middle|\; D \subseteq \binom{[n]}{2} \text{ such that } \big([n], E_G \triangle D\big) \cong H \right\}.$$

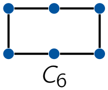
We start by defining several metrics on graphs G, H of the same order $|G| = |H| = n$.
W.l.o.g. $V(G) = V(H) = [n]$.

Graph Edit Distance

$$\delta_{\mathrm{ed}}(G,H) := \min\left\{|D| \;\middle|\; D \subseteq \binom{[n]}{2} \text{ such that } ([n], E_G \triangle D) \cong H\right\}.$$

Examples





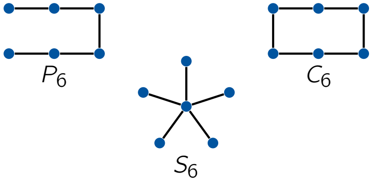
1. P_n := path of length $n-1$; C_n := cycle of length n.
   Then $\delta_{\mathrm{ed}}(P_n, C_n) = 1$.

Edit Distance

We start by defining several metrics on graphs G, H of the same order $|G| = |H| = n$.
W.l.o.g. $V(G) = V(H) = [n]$.

Graph Edit Distance

$$\delta_{\mathrm{ed}}(G,H) := \min\left\{ |D| \;\middle|\; D \subseteq \binom{[n]}{2} \text{ such that } ([n], E_G \triangle D) \cong H \right\}.$$



Examples

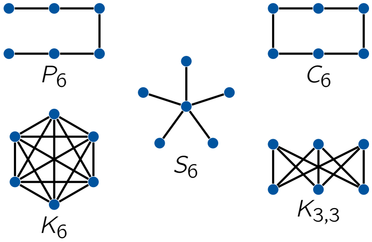
1. P_n := path of length $n-1$; C_n := cycle of length n.
   Then $\delta_{\mathrm{ed}}(P_n, C_n) = 1$.
2. S_n := star with $n-1$ tips.
   Then $\delta_{\mathrm{ed}}(S_n, C_n) = 2n-5$ and $\delta_{\mathrm{ed}}(S_n, P_n) = 2n-6$.

Edit Distance

We start by defining several metrics on graphs G, H of the same order $|G| = |H| = n$.
W.l.o.g. $V(G) = V(H) = [n]$.

Graph Edit Distance

$$\delta_{\mathrm{ed}}(G,H) := \min\left\{ |D| \;\middle|\; D \subseteq \binom{[n]}{2} \text{ such that } ([n], E_G \triangle D) \cong H \right\}.$$



Examples

1. P_n := path of length $n-1$; C_n := cycle of length n.
   Then $\delta_{\mathrm{ed}}(P_n, C_n) = 1$.
2. S_n := star with $n-1$ tips.
   Then $\delta_{\mathrm{ed}}(S_n, C_n) = 2n-5$ and $\delta_{\mathrm{ed}}(S_n, P_n) = 2n-6$.
3. K_n := complete graph on n vertices; $K_{\ell,m}$ complete bipartite graph with parts of size ℓ, m.
   Then $\delta_{\mathrm{ed}}(C_n, K_n) = \frac{n(n-3)}{2}$ and $\delta_{\mathrm{ed}}(K_n, K_{\frac{n}{2},\frac{n}{2}}) = \frac{n}{2}(\frac{n}{2}-1)$ if n is even.

Entrywise Norms

For a matrix $A = (A_{ij}) \in \mathbb{R}^{n \times n}$ and $p \geq 2$ we let

$$\|A\|_{(p)} := \left(\sum_{i,j} |A_{ij}|^p \right)^{\frac{1}{p}}.$$

Entrywise Norms

For a matrix $A = (A_{ij}) \in \mathbb{R}^{n \times n}$ and $p \geq 2$ we let

$$\|A\|_{(p)} := \left(\sum_{i,j} |A_{ij}|^p \right)^{\frac{1}{p}} .$$

$\|A\|_{(2)}$ is known as the Frobenius norm of A.

Entrywise Norms

For a matrix $A = (A_{ij}) \in \mathbb{R}^{n \times n}$ and $p \geq 2$ we let

$$\|A\|_{(p)} := \left(\sum_{i,j} |A_{ij}|^p \right)^{\frac{1}{p}}.$$

$\|A\|_{(2)}$ is known as the Frobenius norm of A.

Observation

$$\delta_{\mathrm{ed}}(G,H) = \frac{1}{2} \min_{\pi \in S_n} \left\| A_G^{\pi} - A_H \right\|_{(1)}$$

where

- *$A_G, A_H \in \{0,1\}^{n \times n}$ are the adjacency matrices of G, H;*
- *S_n is the set of all permutations of $[n]$;*
- *A_G^{π} is the matrix obtained from A_G by simultaneously permuting rows and columns with π.*

Entrywise Norms

For a matrix $A = (A_{ij}) \in \mathbb{R}^{n \times n}$ and $p \geq 2$ we let

$$\|A\|_{(p)} := \left(\sum_{i,j} |A_{ij}|^p \right)^{\frac{1}{p}}.$$

$\|A\|_{(2)}$ is known as the Frobenius norm of A.

Observation

$$\delta_{\text{ed}}(G,H) = \frac{1}{2} \min_{\pi \in S_n} \left\| A_G^\pi - A_H \right\|_{(1)} = \frac{1}{2} \min_{\pi} \left\| A_G^\pi - A_H \right\|_{(2)}^2,$$

where

- $A_G, A_H \in \{0,1\}^{n \times n}$ *are the adjacency matrices of* G, H*;*
- S_n *is the set of all permutations of* $[n]$*;*
- A_G^π *is the matrix obtained from* A_G *by simultaneously permuting rows and columns with* π*.*

Local Edit Distance

$$\delta_{\mathrm{led}}(G,H) := \min\left\{\Delta\big([n],D\big) \;\middle|\; D \subseteq \binom{[n]}{2} \text{ such that } \big([n], E_G \mathbin{\triangle} D\big) \cong H\right\},$$

where $\Delta\big([n],D\big)$ is the maximum degree of the graph $\big([n],D\big)$.

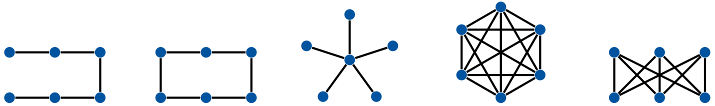
Local Edit Distance

$$\delta_{\mathrm{led}}(G,H) := \min\left\{\Delta([n],D) \;\middle|\; D \subseteq \binom{[n]}{2} \text{ such that } ([n], E_G \,\triangle\, D) \cong H\right\},$$

where $\Delta([n],D)$ is the maximum degree of the graph $([n],D)$.

Examples

1. $\delta_{\mathrm{led}}(P_n, C_n) = 1.$



P_6 C_6 S_6 K_6 $K_{3,3}$

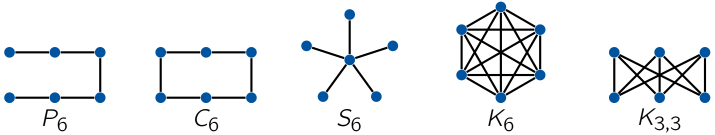
Local Edit Distance

$$\delta_{\text{led}}(G,H) := \min\left\{\Delta([n],D) \;\middle|\; D \subseteq \binom{[n]}{2} \text{ such that } ([n], E_G \triangle D) \cong H\right\},$$

where $\Delta([n],D)$ is the maximum degree of the graph $([n],D)$.

Examples

1. $\delta_{\text{led}}(P_n, C_n) = 1$.
2. $\delta_{\text{led}}(S_n, C_n) = \delta_{\text{led}}(S_n, P_n) = n-3$ for $n \geq 4$.



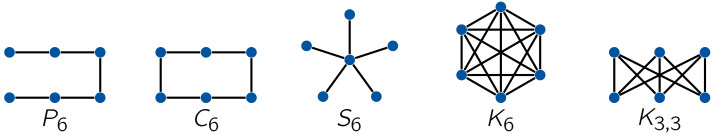
Local Edit Distance

$$\delta_{\text{led}}(G,H) := \min\left\{\Delta([n],D) \;\middle|\; D \subseteq \binom{[n]}{2} \text{ such that } ([n], E_G \mathbin{\triangle} D) \cong H\right\},$$

where $\Delta([n],D)$ is the maximum degree of the graph $([n],D)$.

Examples

1. $\delta_{\text{led}}(P_n, C_n) = 1$.
2. $\delta_{\text{led}}(S_n, C_n) = \delta_{\text{led}}(S_n, P_n) = n-3$ for $n \geq 4$.
3. $\delta_{\text{led}}(C_n, K_n) = n-3$ and $\delta_{\text{led}}(K_n, K_{\frac{n}{2},\frac{n}{2}}) = \frac{n}{2} - 1$ if n is even.



Operator Norms

For a matrix $A = (A_{ij}) \in \mathbb{R}^{n \times n}$ and $p \geq 2$ we let

$$\|A\|_p := \max_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p},$$

Operator Norms

For a matrix $A = (A_{ij}) \in \mathbb{R}^{n \times n}$ and $p \geq 2$ we let

$$\|A\|_p := \max_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p},$$

where for a vector $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$, by $\|\mathbf{a}\|_p = \left(\sum_{i=1}^{n} |a_i|^p\right)^{\frac{1}{p}}$ we denote the usual ℓ_p vector norm.

Operator Norms

For a matrix $A = (A_{ij}) \in \mathbb{R}^{n \times n}$ and $p \geq 2$ we let

$$\|A\|_p := \max_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p},$$

where for a vector $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$, by $\|\mathbf{a}\|_p = \left(\sum_{i=1}^n |a_i|^p\right)^{\frac{1}{p}}$ we denote the usual ℓ_p vector norm.

$\|A\|_2$ is known as the spectral norm of A.

Operator Norms

For a matrix $A = (A_{ij}) \in \mathbb{R}^{n \times n}$ and $p \geq 2$ we let

$$\|A\|_p := \max_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p},$$

where for a vector $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$, by $\|\mathbf{a}\|_p = \left(\sum_{i=1}^n |a_i|^p\right)^{\frac{1}{p}}$ we denote the usual ℓ_p vector norm.

$\|A\|_2$ is known as the spectral norm of A.

Observation (Gervens, G. 2022)

$$\delta_{\text{led}}(G, H) = \min_{\pi \in S_n} \left\| A_G^{\pi} - A_H \right\|_1.$$

Operator Norms

For a matrix $A = (A_{ij}) \in \mathbb{R}^{n \times n}$ and $p \geq 2$ we let

$$\|A\|_p := \max_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p},$$

where for a vector $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$, by $\|\mathbf{a}\|_p = \left(\sum_{i=1}^n |a_i|^p\right)^{\frac{1}{p}}$ we denote the usual ℓ_p vector norm.

$\|A\|_2$ is known as the spectral norm of A.

Observation (Gervens, G. 2022)

$$\delta_{\mathrm{led}}(G, H) = \min_{\pi \in S_n} \left\|A_G^\pi - A_H\right\|_1.$$

Remark

Another interesting graph metric is derived from the spectral norm:

$$\delta_{\mathrm{sp}}(G, H) := \min_{\pi \in S_n} \left\|A_G^\pi - A_H\right\|_2.$$

For sets $S, T \subseteq V(G)$:

$e_G(S,T) :=$ number of edges of G between S and T, counting edges with both endvertices in $S \cap T$ twice.

For sets $S, T \subseteq V(G)$:

$e_G(S,T) :=$ number of edges of G between S and T, counting edges with both endvertices in $S \cap T$ twice.

Cut Distance

$$\delta_\square(G,H) := \min_{\pi \in S_n} \max_{S,T \subseteq [n]} \Big| e_G(S,T) - e_H\big(\pi(S), \pi(T)\big) \Big|.$$

For sets $S, T \subseteq V(G)$:

$$e_G(S,T) := \text{number of edges of } G \text{ between } S \text{ and } T\text{, counting edges with both endvertices in } S \cap T \text{ twice.}$$

Cut Distance

$$\delta_\square(G,H) := \min_{\pi \in S_n} \max_{S,T \subseteq [n]} \Big| e_G(S,T) - e_H\big(\pi(S), \pi(T)\big) \Big|.$$

Theorem (Lovász (?))

For random graphs $G, H \sim \mathcal{G}(n, \frac{1}{2})$, with high probability it holds that

$$\delta_\square(G,H) = O(n^{3/2}) \quad \textit{and} \quad \delta_{\mathrm{ed}}(G,H) = \Omega(n^2).$$

Cut Norm

For a matrix $A = (A_{ij}) \in \mathbb{R}^{n \times n}$ we let

$$\|A\|_{\square} := \max_{S,T \subseteq [n]} \left| \sum_{i \in S, j \in T} A(i,j) \right|.$$

Cut Norm

For a matrix $A = (A_{ij}) \in \mathbb{R}^{n \times n}$ we let

$$\|A\|_\square := \max_{S,T \subseteq [n]} \left| \sum_{i \in S, j \in T} A(i,j) \right|.$$

Observation

$$\delta_\square(G,H) = \min_{\pi \in S_n} \left\| A_G^\pi - A_H \right\|_\square.$$

Local Edit Distance via Matrix Norms

Cut Norm

For a matrix $A = (A_{ij}) \in \mathbb{R}^{n \times n}$ we let

$$\|A\|_\square := \max_{S,T \subseteq [n]} \left| \sum_{i \in S, j \in T} A(i,j) \right|.$$

Observation

$$\delta_\square(G,H) = \min_{\pi \in S_n} \left\| A_G^\pi - A_H \right\|_\square.$$

Theorem (Nikiforov 2009)

$$\frac{1}{n}\delta_\square(G,H) \leq \delta_{\mathrm{sp}}(G,H) \leq 2\sqrt{\delta_\square(G,H)}.$$

Observation

The metrics scale differently. Denoting the edgeless graph with n vertices by L_n, for all metrics δ considered so far we have

$$\delta(L_n, K_n) = \max\{\delta(G, H) \mid |G| = |H| = n\}.$$

Observation

The metrics scale differently. Denoting the edgeless graph with n vertices by L_n, for all metrics δ considered so far we have

$$\delta(L_n, K_n) = \max\{\delta(G, H) \mid |G| = |H| = n\}.$$

Furthermore,

$$2\delta_{\mathrm{ed}}(L_n, K_n) = \delta_{\square}(G, H) = n(n-1) \sim n^2,$$
$$\delta_{\mathrm{led}}(L_n, K_n) = \delta_{\mathrm{sp}}(G, H) = n - 1 \sim n.$$

Observation

The metrics scale differently. Denoting the edgeless graph with n vertices by L_n, for all metrics δ considered so far we have

$$\delta(L_n, K_n) = \max\{\delta(G,H) \mid |G| = |H| = n\}.$$

Furthermore,

$$2\delta_{\mathrm{ed}}(L_n, K_n) = \delta_{\square}(G,H) = n(n-1) \sim n^2,$$
$$\delta_{\mathrm{led}}(L_n, K_n) = \delta_{\mathrm{sp}}(G,H) = n-1 \sim n.$$

Normalised Metrics

$$\widehat{\delta_{\mathrm{ed}}}(G,H) := \frac{2}{n^2}\delta_{\mathrm{ed}}(G,H), \qquad \widehat{\delta_{\square}}(G,H) := \frac{1}{n^2}\delta(G,H),$$
$$\widehat{\delta_{\mathrm{led}}}(G,H) := \frac{1}{n}\delta_{\mathrm{ed}}(G,H), \qquad \widehat{\delta_{\mathrm{sp}}}(G,H) := \frac{1}{n}\delta_{\mathrm{sp}}(G,H).$$

Example

$$\widehat{\delta_{\mathrm{ed}}}(S_n, C_n) = \frac{2(2n-5)}{n^2} \xrightarrow[n\to\infty]{} 0 \qquad \text{and} \qquad \widehat{\delta_{\mathrm{led}}}(S_n, C_n) = \frac{n-3}{n} \xrightarrow[n\to\infty]{} 1.$$

Example

$$\widehat{\delta_{\mathrm{ed}}}(S_n, C_n) = \frac{2(2n-5)}{n^2} \xrightarrow[n\to\infty]{} 0 \qquad \text{and} \qquad \widehat{\delta_{\mathrm{led}}}(S_n, C_n) = \frac{n-3}{n} \xrightarrow[n\to\infty]{} 1.$$

Corollary

For random graphs $G, H \sim \mathscr{G}(n, \frac{1}{2})$, with high probability it holds that

$$\widehat{\delta_{\square}}(G,H) \xrightarrow[n\to\infty]{} 0 \qquad \textit{and} \qquad \widehat{\delta_{\mathrm{ed}}}(G,H) = \Omega(1).$$

Relations Between the Normalised Metrics

Example

$$\widehat{\delta_{\mathrm{ed}}}(S_n, C_n) = \frac{2(2n-5)}{n^2} \xrightarrow[n\to\infty]{} 0 \qquad \text{and} \qquad \widehat{\delta_{\mathrm{led}}}(S_n, C_n) = \frac{n-3}{n} \xrightarrow[n\to\infty]{} 1.$$

Corollary

For random graphs $G, H \sim \mathscr{G}(n, \frac{1}{2})$, with high probability it holds that

$$\widehat{\delta_{\square}}(G,H) \xrightarrow[n\to\infty]{} 0 \qquad \textit{and} \qquad \widehat{\delta_{\mathrm{ed}}}(G,H) = \Omega(1).$$

Corollary

$$\widehat{\delta_{\square}}(G,H) \leq \widehat{\delta_{\mathrm{sp}}}(G,H) \leq 2\sqrt{\widehat{\delta_{\square}}(G,H)}.$$

The Blow-up Construction

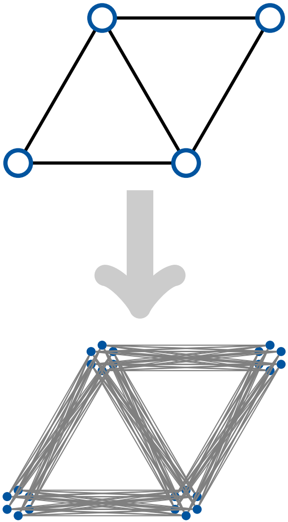
$G^{\bullet k}$ obtained from G by

- replacing each node by a set of k nodes;
- replacing each edge by a complete bipartite graph.

The Blow-up Construction

$G^{\text{💣}k}$ obtained from G by

- replacing each node by a set of k nodes;
- replacing each edge by a complete bipartite graph.

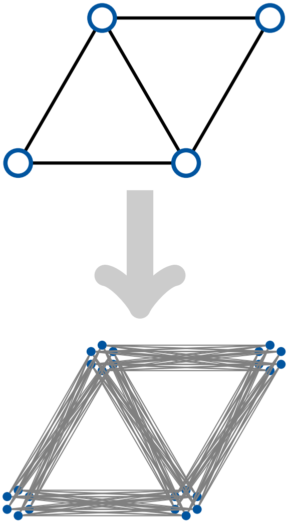


Comparing Graphs of Distinct Sizes

The Blow-up Construction

$G^{\text{💣}k}$ obtained from G by

- replacing each node by a set of k nodes;
- replacing each edge by a complete bipartite graph.



Blow-up Distance

For every normalised metric $\widetilde{\delta}$ and graphs G, H of order $m := |G|, n := |H|$ we define

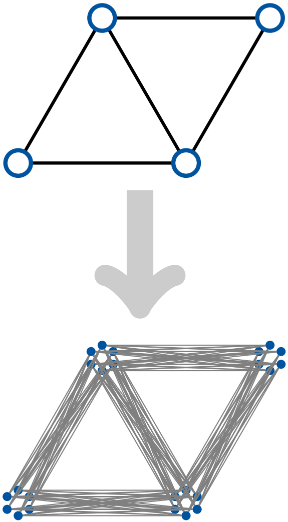
$$\delta^{\text{💣}}(G,H) := \lim_{\ell \to \infty} \widetilde{\delta}(G^{\text{💣}n\ell}, H^{\text{💣}m\ell}).$$

Comparing Graphs of Distinct Sizes

The Blow-up Construction

$G^{\text{💣}k}$ obtained from G by

- replacing each node by a set of k nodes;
- replacing each edge by a complete bipartite graph.



Blow-up Distance

For every normalised metric $\widetilde{\delta}$ and graphs G, H of order $m := |G|, n := |H|$ we define

$$\delta^{\text{💣}}(G, H) := \lim_{\ell \to \infty} \widetilde{\delta}(G^{\text{💣}n\ell}, H^{\text{💣}m\ell}).$$

Remark

There is an alternative way of defining the blow-up distances through an optimal transport formulation.

Observation

Let δ be one of the metrics considered so far.

1. *For all graphs G and $k \geq 1$ we have $\delta^{\text{💣}}(G, G^{\text{💣}k}) = 0$.*

Properties of the Blow-up Distances

Observation

Let δ be one of the metrics considered so far.

1. *For all graphs G and $k \geq 1$ we have $\delta^{💣}(G, G^{💣k}) = 0$.*
2. *For all graphs G, H of the same order it holds that*

$$\delta^{💣}(G, H) \leq \widehat{\delta}(G, H).$$

Furthermore, there are example of graphs where the inequality is strict.

Properties of the Blow-up Distances

Observation

Let δ be one of the metrics considered so far.

1. *For all graphs G and $k \geq 1$ we have $\delta^{\bullet}(G, G^{\bullet k}) = 0$.*
2. *For all graphs G, H of the same order it holds that*

$$\delta^{\bullet}(G,H) \leq \widehat{\delta}(G,H).$$

Furthermore, there are example of graphs where the inequality is strict.

Theorem (Borgs et al. 2008, Pikhurko 2010)

For all graphs G, H of the same order we have

$$\delta^{\bullet}_{\mathrm{ed}}(G,H) \geq \frac{1}{3}\widehat{\delta_{\mathrm{ed}}}(G,H) \qquad \textit{and} \qquad \delta^{\bullet}_{\square}(G,H) \geq \left(\frac{\widehat{\delta_{\square}}(G,H)}{32}\right)^{67}.$$

A vector embedding (of graphs) is an isomorphism invariant mapping $\eta : \mathcal{G} \to \mathbb{H}$ from graphs to some Hilbert space $\mathbb{H}$ (called the latent space).

A vector embedding (of graphs) is an isomorphism invariant mapping $\eta : \mathcal{G} \to \mathbb{H}$ from graphs to some Hilbert space $\mathbb{H}$ (called the latent space).

A (possibly infinite-dimensional) vector space with an inner product that is complete w.r.t. the metric defined by the inner product.

A vector embedding (of graphs) is an isomorphism invariant mapping $\eta : \mathcal{G} \to \mathbb{H}$ from graphs to some Hilbert space $\mathbb{H}$ (called the latent space).

Every vector embedding $\eta : \mathcal{G} \to \mathbb{H}$ defines a graph metric:

$$\delta_\eta(G, H) := \|\eta(G) - \eta(H)\|_{\mathbb{H}} = \sqrt{\left\langle \eta(G) - \eta(H), \eta(G) - \eta(H) \right\rangle_{\mathbb{H}}}.$$

A vector embedding (of graphs) is an isomorphism invariant mapping $\eta : \mathcal{G} \to \mathbb{H}$ from graphs to some Hilbert space $\mathbb{H}$ (called the latent space).

Every vector embedding $\eta : \mathcal{G} \to \mathbb{H}$ defines a graph metric:

$$\delta_\eta(G, H) := \|\eta(G) - \eta(H)\|_{\mathbb{H}} = \sqrt{\Big\langle \eta(G) - \eta(H), \eta(G) - \eta(H) \Big\rangle_{\mathbb{H}}}.$$

Example

$\eta : \mathcal{G} \to \mathbb{R}^k$ mapping G to the k largest eigenvalues of its adjacency matrix A_G in decreasing order.

A vector embedding (of graphs) is an isomorphism invariant mapping $\eta : \mathscr{G} \to \mathbb{H}$ from graphs to some Hilbert space $\mathbb{H}$ (called the latent space).

Every vector embedding $\eta : \mathscr{G} \to \mathbb{H}$ defines a graph metric:

$$\delta_\eta(G,H) := \|\eta(G) - \eta(H)\|_{\mathbb{H}} = \sqrt{\big\langle \eta(G) - \eta(H), \eta(G) - \eta(H) \big\rangle_{\mathbb{H}}}.$$

Example

$\eta : \mathscr{G} \to \mathbb{R}^k$ mapping G to the k largest eigenvalues of its adjacency matrix A_G in decreasing order.

Idea

Define vector embedding in such a way that the geometry of the latent space has a semantic meaning on the space of graphs.

Example (Subgraph Embeddings)

For graphs $F_1, \ldots, F_k$, we define a vector embedding $\mathsf{sub}_{F_1,\ldots,F_k} : \mathscr{G} \to \mathbb{R}^k$ by

$$\mathsf{Sub}_{F_1,\ldots,F_k}(G) := \big(\mathsf{sub}(F_1, G), \ldots, \mathsf{sub}(F_k, G)\big),$$

where $\mathsf{sub}(F, G)$ is the number of subgraphs of G isomorphic to F.

Example (Subgraph Embeddings)

For graphs $F_1,\dots,F_k$, we define a vector embedding $\mathsf{sub}_{F_1,\dots,F_k} : \mathscr{G} \to \mathbb{R}^k$ by

$$\mathsf{Sub}_{F_1,\dots,F_k}(G) := \big(\mathsf{sub}(F_1, G),\dots,\mathsf{sub}(F_k, G)\big),$$

where $\mathsf{sub}(F, G)$ is the number of subgraphs of G isomorphic to F.

Graphlet kernels are based on this idea.

Example (Subgraph Embeddings)

For graphs $F_1, \dots, F_k$, we define a vector embedding $\mathbf{sub}_{F_1,\dots,F_k} : \mathcal{G} \to \mathbb{R}^k$ by

$$\mathsf{Sub}_{F_1,\dots,F_k}(G) := \big(\mathsf{sub}(F_1, G), \dots, \mathsf{sub}(F_k, G)\big),$$

where $\mathsf{sub}(F, G)$ is the number of subgraphs of G isomorphic to F.

Graphlet kernels are based on this idea.

Example (Logic Embeddings)

For formulas $\varphi_1, \dots, \varphi_k$ of some logic on graphs (say, first-order logic), we define a vector embedding $\mathbf{mod}_{\varphi_1,\dots,\varphi_k} : \mathcal{G} \to \mathbb{R}^k$ by

$$\mathsf{Mod}_{\varphi_1,\dots,\varphi_k}(G) := (b_1, \dots, b_k),$$

where $b_i = 1$ if $G \models \varphi_i$ and $b_i = 0$ otherwise.

A graph kernel is a function $\kappa : \mathcal{G} \times \mathcal{G} \to \mathbb{R}$ such that

- $\kappa(G, H) = \kappa(H, G)$ for all graphs G, H;
- for all graphs $G_1, \ldots, G_n$, the $(n \times n)$-matrix K with entries $K_{ij} := \kappa(G_i, G_j)$ is positive semi-definite.

A graph kernel is a function $\kappa : \mathscr{G} \times \mathscr{G} \to \mathbb{R}$ such that

- $\kappa(G,H) = \kappa(H,G)$ for all graphs G,H;
- for all graphs $G_1,\ldots,G_n$, the $(n \times n)$-matrix K with entries $K_{ij} := \kappa(G_i,G_j)$ is positive semi-definite.

Theorem (Folklore)

Let κ be a graph kernel. Then there is a vector embedding $\eta : \mathscr{G} \to \mathbb{H}$ such that

$$\kappa(G,H)) = \Big\langle \eta(G), \eta(H) \Big\rangle_{\mathbb{H}}$$

for all graph G,H.

A graph kernel is a function $\kappa : \mathcal{G} \times \mathcal{G} \to \mathbb{R}$ such that

- $\kappa(G,H) = \kappa(H,G)$ for all graphs G,H;
- for all graphs $G_1,\dots,G_n$, the $(n \times n)$-matrix K with entries $K_{ij} := \kappa(G_i,G_j)$ is positive semi-definite.

Theorem (Folklore)

Let κ be a graph kernel. Then there is a vector embedding $\eta : \mathcal{G} \to \mathbb{H}$ such that

$$\kappa(G,H)) = \left\langle \eta(G), \eta(H) \right\rangle_{\mathbb{H}}$$

for all graph G,H.

Example

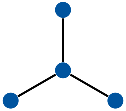
Weisfeiler-Leman kernels (Shervashidze et al., 2009) are efficiently computable without ever explicitly computing the vector embedding into the infinite dimensional latent space.

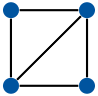
Counting Homomorphisms

A homomorphism between graphs is a mapping between the vertices that preserves adjacency.

A homomorphism between graphs is a mapping between the vertices that preserves adjacency.

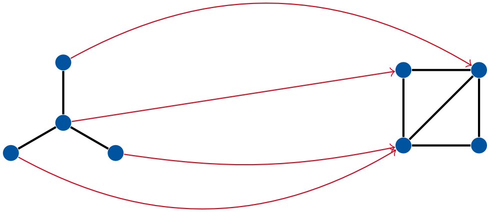
Example





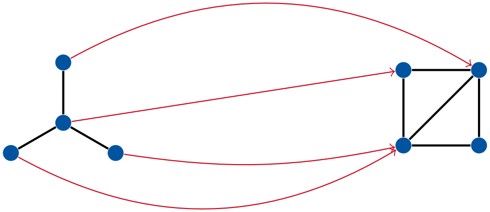
A homomorphism between graphs is a mapping between the vertices that preserves adjacency.

Example



A homomorphism between graphs is a mapping between the vertices that preserves adjacency.

Example

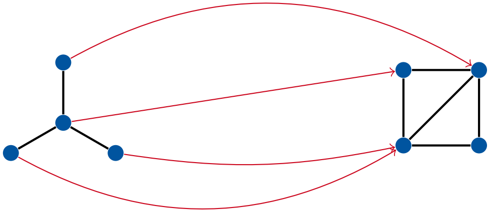


For all graphs F, G of orders $k := |F|, n := |G|$:

$$\hom(F, G) = \text{number of homomorphisms from } F \text{ to } G$$

A homomorphism between graphs is a mapping between the vertices that preserves adjacency.

Example



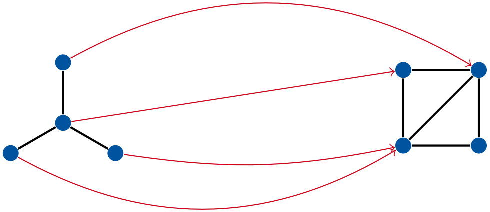
For all graphs F, G of orders $k := |F|, n := |G|$:

$$\mathrm{hom}(F,G) = \text{number of homomorphisms from } F \text{ to } G$$

$$\mathrm{hd}(F,G) = \frac{1}{n^k}\,\mathrm{hom}(F,G)$$

A homomorphism between graphs is a mapping between the vertices that preserves adjacency.

Example



For all graphs F, G of orders $k := |F|, n := |G|$:

$$\mathrm{hom}(F, G) = \text{number of homomorphisms from } F \text{ to } G$$

$$\mathrm{hd}(F, G) = \frac{1}{n^k} \mathrm{hom}(F, G) = \Pr_h(h \text{ is a homomorphism from } F \text{ to } G),$$

where h is mapping from $V(F)$ to $V(G)$ chosen uniformly at random.

For every class $\mathscr{F}$ of graphs, we define a vector embedding $\mathsf{Hd}_{\mathscr{F}} : \mathscr{G} \to \mathbb{R}^{\mathscr{F}}$ by

$$\mathsf{Hd}_{\mathscr{F}}(G) := \big(\mathsf{hd}(F, G)\big)_{F \in \mathscr{F}}.$$

For every class $\mathscr{F}$ of graphs, we define a vector embedding $\mathsf{Hd}_{\mathscr{F}} : \mathscr{G} \to \mathbb{R}^{\mathscr{F}}$ by

$$\mathsf{Hd}_{\mathscr{F}}(G) := \big(\mathsf{hd}(F, G)\big)_{F \in \mathscr{F}}.$$

We can define an inner product on $\mathbb{R}^{\mathscr{F}}$ (even for infinite $\mathscr{F}$) and turn it into a Hilbert space.

For every class $\mathscr{F}$ of graphs, we define a vector embedding $\mathbf{Hd}_{\mathscr{F}} : \mathscr{G} \to \mathbb{R}^{\mathscr{F}}$ by

$$\mathbf{Hd}_{\mathscr{F}}(G) := \big(\mathrm{hd}(F,G)\big)_{F \in \mathscr{F}}.$$

We can define an inner product on $\mathbb{R}^{\mathscr{F}}$ (even for infinite $\mathscr{F}$) and turn it into a Hilbert space.

It gives us the (normalised) graph metric

$$\delta_{\mathscr{F}}(G,H) = \sqrt{\sum_{k \in \mathbb{N}} \frac{1}{2^k |\mathscr{F}_k|} \sum_{F \in \mathscr{F}_k} \big(\mathrm{hd}(F,G) - \mathrm{hd}(F,H)\big)^2},$$

where $\mathscr{F}_k$ is the set of graphs of order k in $\mathscr{F}$.

Operational View

Two graphs are similar if they can easily be transformed into each other.

Examples

edit distance, all distances based on matrix norms

Advantage:

gives an alignment between graphs

Two Sides of Similarity

Operational View

Two graphs are similar if they can easily be transformed into each other.

Examples

edit distance, all distances based on matrix norms

Advantage:

gives an alignment between graphs

Declarative View

Two graphs are similar if they have similar properties.

Examples

distances based on vector embeddings

Advantage:

can be adapted to specific targets by selecting properties/features

Operational View

Two graphs are similar if they can easily be transformed into each other.

Examples

edit distance, all distances based on matrix norms

Advantage:

gives an alignment between graphs

Declarative View

Two graphs are similar if they have similar properties.

Examples

distances based on vector embeddings

Advantage:

can be adapted to specific targets by selecting properties/features

Question: How do these relate?

Comparing the Topologies (Definitions)

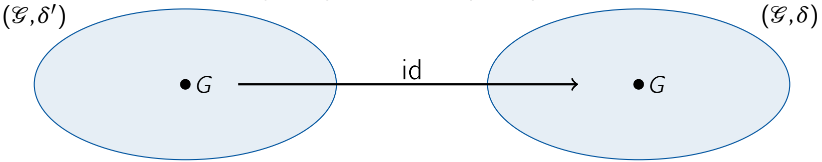
Let δ, δ' be graph metrics. We say that δ is coarser than δ' (we write $\delta \sqsubseteq \delta'$) if for all $\epsilon > 0$ there is an $\epsilon' > 0$ such that

$$\delta'(G, H) < \epsilon' \implies \delta(G, H) < \epsilon.$$

Comparing the Topologies (Definitions)

Let δ, δ' be graph metrics. We say that δ is coarser than δ' (we write $\delta \sqsubseteq \delta'$) if for all $\epsilon > 0$ there is an $\epsilon' > 0$ such that

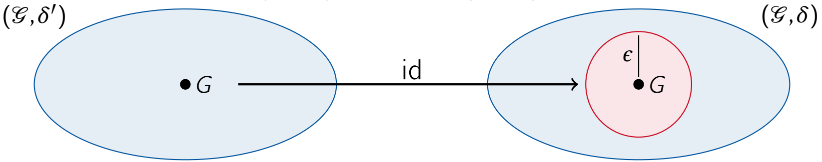
$$\delta'(G,H) < \epsilon' \implies \delta(G,H) < \epsilon.$$



Comparing the Topologies (Definitions)

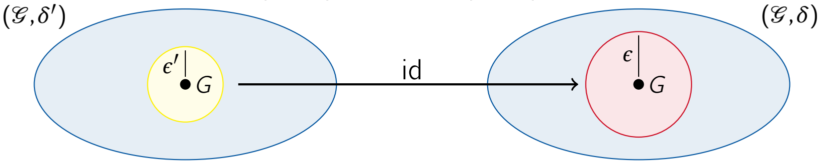
Let δ, δ' be graph metrics. We say that δ is coarser than δ' (we write $\delta \sqsubseteq \delta'$) if for all $\epsilon > 0$ there is an $\epsilon' > 0$ such that

$$\delta'(G,H) < \epsilon' \implies \delta(G,H) < \epsilon.$$



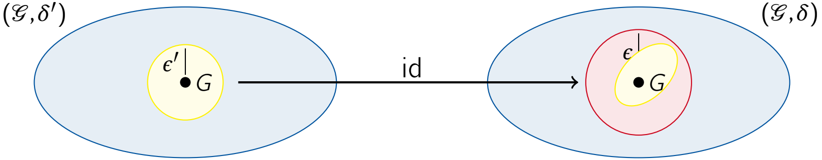
Let δ, δ' be graph metrics. We say that δ is coarser than δ' (we write $\delta \sqsubseteq \delta'$) if for all $\epsilon > 0$ there is an $\epsilon' > 0$ such that

$$\delta'(G,H) < \epsilon' \implies \delta(G,H) < \epsilon.$$



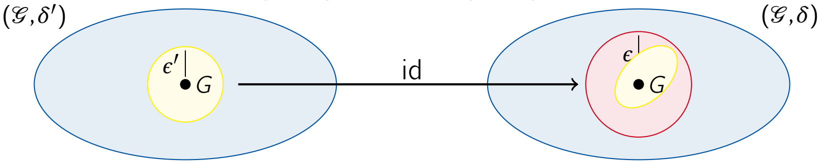
Let δ, δ' be graph metrics. We say that δ is coarser than δ' (we write $\delta \sqsubseteq \delta'$) if for all $\epsilon > 0$ there is an $\epsilon' > 0$ such that

$$\delta'(G,H) < \epsilon' \implies \delta(G,H) < \epsilon.$$



Let δ, δ' be graph metrics. We say that δ is coarser than δ' (we write $\delta \sqsubseteq \delta'$) if for all $\epsilon > 0$ there is an $\epsilon' > 0$ such that

$$\delta'(G,H) < \epsilon' \implies \delta(G,H) < \epsilon.$$

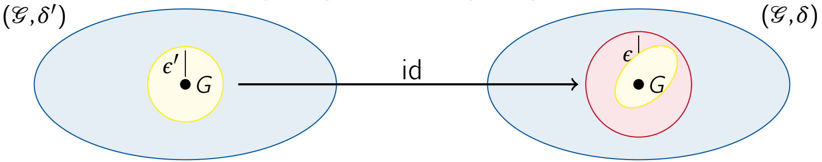


Moreover, δ and δ' are topologically equivalent (we write $\delta' \equiv \delta$) if $\delta \sqsubseteq \delta'$ and $\delta' \sqsubseteq \delta$, and δ is strictly coarser than δ' (we write $\delta \sqsubset \delta'$) if $\delta \sqsubseteq \delta'$ and $\delta' \not\sqsubseteq \delta$.

Comparing the Topologies (Definitions)

Let δ, δ' be graph metrics. We say that δ is coarser than δ' (we write $\delta \sqsubseteq \delta'$) if for all $\epsilon > 0$ there is an $\epsilon' > 0$ such that

$$\delta'(G,H) < \epsilon' \implies \delta(G,H) < \epsilon.$$



Moreover, δ and δ' are topologically equivalent (we write $\delta' \equiv \delta$) if $\delta \sqsubseteq \delta'$ and $\delta' \sqsubseteq \delta$, and δ is strictly coarser than δ' (we write $\delta \sqsubset \delta'$) if $\delta \sqsubseteq \delta'$ and $\delta' \not\sqsubseteq \delta$.

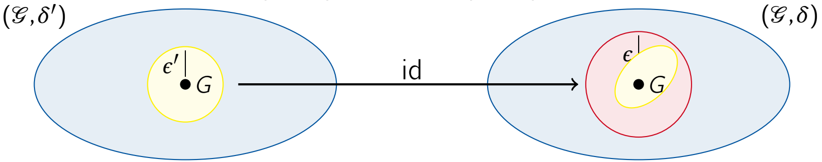
Observation

$\delta \sqsubseteq \delta'$ if and only if the identity is a continuous mapping from the metric space $(\mathcal{G}, \delta')$ to the metric space $(\mathcal{G}, \delta)$.

Comparing the Topologies (Definitions)

Let δ, δ' be graph metrics. We say that δ is coarser than δ' (we write $\delta \sqsubseteq \delta'$) if for all $\epsilon > 0$ there is an $\epsilon' > 0$ such that

$$\delta'(G,H) < \epsilon' \implies \delta(G,H) < \epsilon.$$



Moreover, δ and δ' are topologically equivalent (we write $\delta' \equiv \delta$) if $\delta \sqsubseteq \delta'$ and $\delta' \sqsubseteq \delta$, and δ is strictly coarser than δ' (we write $\delta \sqsubset \delta'$) if $\delta \sqsubseteq \delta'$ and $\delta' \not\sqsubseteq \delta$.

Observation

$\delta \sqsubseteq \delta'$ if and only if the identity is a continuous mapping from the metric space $(\mathscr{G}, \delta')$ to the metric space $(\mathscr{G}, \delta)$.

Moreover, $\delta \equiv \delta'$ if and only if δ and δ' define the same topology on $\mathscr{G}$.

Comparing the Topologies (Results)

Graph Metrics Based on Matrix Norms

$$\delta^{\Bomb}_{\square} \equiv \delta^{\Bomb}_{\mathrm{sp}} \sqsubset \delta^{\Bomb}_{\mathrm{ed}} \sqsubset \delta^{\Bomb}_{\mathrm{led}}.$$

Comparing the Topologies (Results)

Graph Metrics Based on Matrix Norms

$$\delta^{\faBomb}_{\square} \equiv \delta^{\faBomb}_{\mathsf{sp}} \sqsubset \delta^{\faBomb}_{\mathsf{ed}} \sqsubset \delta^{\faBomb}_{\mathsf{led}}.$$

Theorem (Borgs et al. 2008)

$$\delta^{\faBomb}_{\square} \equiv \delta_{\mathscr{G}}.$$

Theorem (G. et al. 2018, G. and Gervens 2022)

Computing δ_{ed},δ_{led},δ_{sp} is NP-complete even if both input graphs are trees.

Theorem (G. et al. 2018, G. and Gervens 2022)

Computing $\delta_{\mathrm{ed}}, \delta_{\mathrm{led}}, \delta_{\mathrm{sp}}$ is NP-complete even if both input graphs are trees.

Remarks

- Various other hardness results are known. In particular, the distances are also hard to approximate.

Theorem (G. et al. 2018, G. and Gervens 2022)

Computing $\delta_{\mathrm{ed}}, \delta_{\mathrm{led}}, \delta_{\mathrm{sp}}$ *is NP-complete even if both input graphs are trees.*

Remarks

- Various other hardness results are known. In particular, the distances are also hard to approximate.
- The problem of computing the distances is closely related to the notoriously hard quadratic assignment problem from combinatorial optimisation.

For graph metrics δ defined by matrix norms $\|\cdot\|$:

$$\delta(G,H) = \min_{\pi \in S_n} \|A^\pi - B\|$$

,

For graph metrics δ defined by matrix norms $\|\cdot\|$:

$$\begin{aligned}\delta(G,H) &= \min_{\pi \in S_n} \|A^{\pi} - B\| \\ &= \min_{P} \|AP - PB\|,\end{aligned}$$

where min ranges over all $(n \times n)$ permutation matrices P.

For graph metrics δ defined by matrix norms $\|\cdot\|$:

$$\begin{aligned}\delta(G,H) &= \min_{\pi \in S_n} \|A^{\pi} - B\| \\ &= \min_{P} \|AP - PB\|,\end{aligned}$$

where min ranges over all $(n \times n)$ permutation matrices P.

Convex Relaxation

$$\delta^*(G,H) = \min_{X} \|AX - XB\|,$$

where min ranges over all $(n \times n)$ doubly stochastic matrices X.

For graph metrics δ defined by matrix norms $\|\cdot\|$:

$$\begin{aligned}\delta(G,H) &= \min_{\pi \in S_n} \|A^{\pi} - B\| \\ &= \min_{P} \|AP - PB\|,\end{aligned}$$

where min ranges over all $(n \times n)$ permutation matrices P.

Convex Relaxation

$$\delta^*(G,H) = \min_{X} \|AX - XB\|,$$

where min ranges over all $(n \times n)$ doubly stochastic matrices X.

Theorem

δ^*_{ed}, δ^*_{led}, δ^*_{sp} *can be computed in polynomial time.*

$\mathscr{T}$ = class of all trees

$\mathscr{T}$ = class of all trees

Theorem (Böker 2021, Grebík and Rocha 2022)

$$\delta_{\mathrm{sp}}^{*} \quad \equiv \quad \delta_{\square}^{*} \quad \equiv \quad \delta_{\mathscr{T}}\,.$$

Learned Embeddings

Graph neural networks are deep learning architectures for graphs. We can use them to learn vector embeddings of graphs from data.

Graph neural networks are deep learning architectures for graphs. We can use them to learn vector embeddings of graphs from data.

Theorem (Böker et al. 2023)

1. *Let* δ *be a bounded graph metric computed by a GNN. Then* $\delta \sqsubseteq \delta_{\mathscr{T}}$.

Graph neural networks are deep learning architectures for graphs. We can use them to learn vector embeddings of graphs from data.

Theorem (Böker et al. 2023)

1. *Let δ be a bounded graph metric computed by a GNN. Then $\delta \sqsubseteq \delta_{\mathscr{T}}$.*
2. *For every n there is a graph metric $\delta^{(n)}$ computed by a GNN such that $\delta^{(n)} \equiv \delta_{\mathscr{T}}$ on graphs of order at most n.*

- Matrix norms and homomorphism counts give principled ways to define graph metrics, and there are deep and interesting connections between them.

Concluding Remarks

- Matrix norms and homomorphism counts give principled ways to define graph metrics, and there are deep and interesting connections between them.
- There are other natural ways of defining graph metrics, most notably by viewing graphs as finite metric spaces and using notions from discrete geometry (e.g. Gromov-Wasserstein metric).

Concluding Remarks

- Matrix norms and homomorphism counts give principled ways to define graph metrics, and there are deep and interesting connections between them.
- There are other natural ways of defining graph metrics, most notably by viewing graphs as finite metric spaces and using notions from discrete geometry (e.g. Gromov-Wasserstein metric).
- The normalisation we chose is tailored towards dense graphs. Under our normalised metrics, all sparse graphs are close to the empty graph.

- Matrix norms and homomorphism counts give principled ways to define graph metrics, and there are deep and interesting connections between them.
- There are other natural ways of defining graph metrics, most notably by viewing graphs as finite metric spaces and using notions from discrete geometry (e.g. Gromov-Wasserstein metric).
- The normalisation we chose is tailored towards dense graphs. Under our normalised metrics, all sparse graphs are close to the empty graph.

  There is no fully developed theory for sparse graphs.

Further Reading

Most of the material of this talk is covered in the following article.

Grohe, M. (2024). Some Thoughts on Graph Similarity.
arXiv:2411.10182