UNIVERSITÄT BONN

Wasserstein Weisfeiler-Lehman Graph Kernels

Matteo Togninalli, Felipe Llinares-López, Elisabetta Ghisu, Bastian Rieck, Karsten Borgwardt – ETH ZURICH

Fabrice Beaumont

Rheinische Friedrich-Wilhelms-Universität Bonn

01.07.2020

UNIVERSITÄT BONN

Wasserstein Weisfeiler-Lehman

Graph Kernels

Matteo Togninalli, Felipe Llinares-López, Elisabetta Ghisu, Bastian Rieck, Karsten Borgwardt – ETH ZURICH

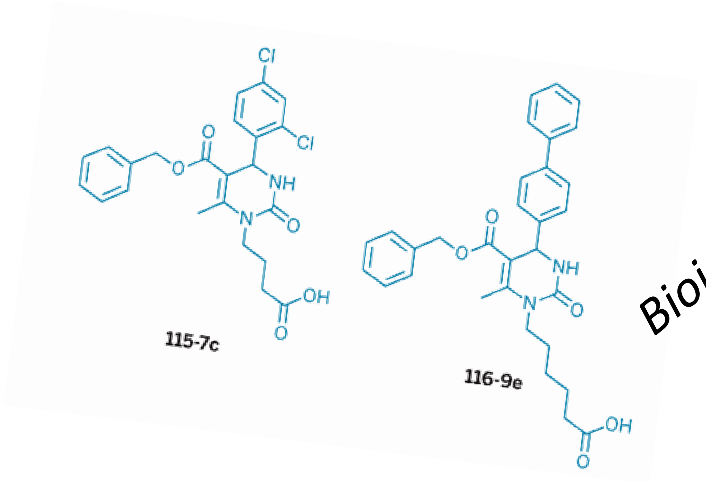




01.07.2020

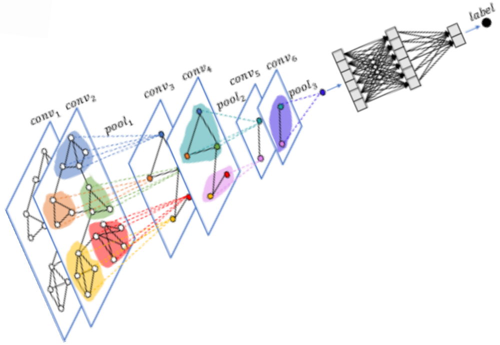
Motivation

TASK: Graph classification (via graph kernels as similarity measure)

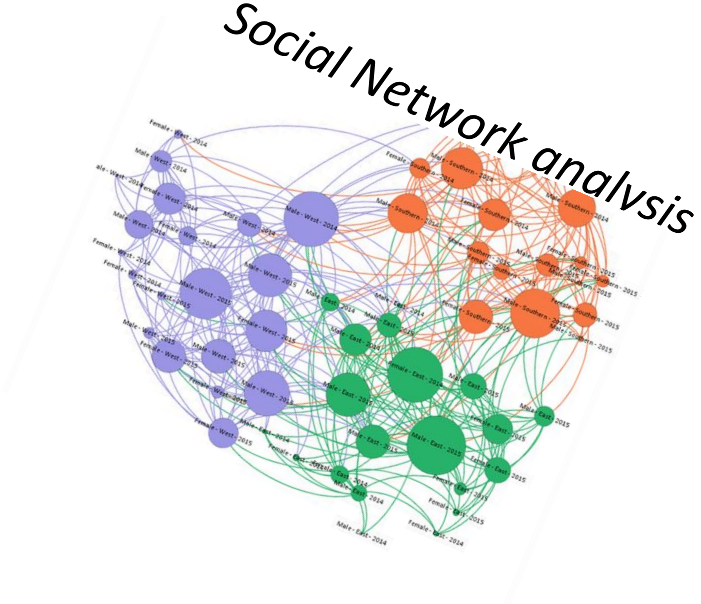


Bioinformatics

Computer vision



Social Network analysis



Motivation

TASK: Graph classification (via graph kernels as similarity measure)

Solution: $\mathfrak{R}$-Convolution kernels

Problem: **1.** Naive **aggregation** of **substructures** may disregard valuable information

2. Only a few approaches extendible for **continuously labelled** graphs

Motivation

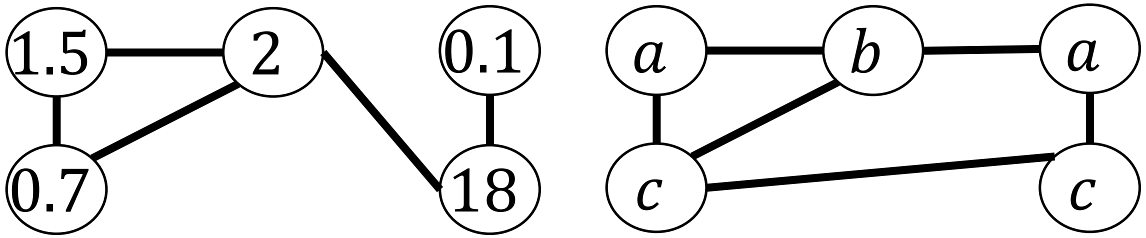
UNIVERSITÄT BONN

TASK: Graph classification (via graph kernels as similarity measure)

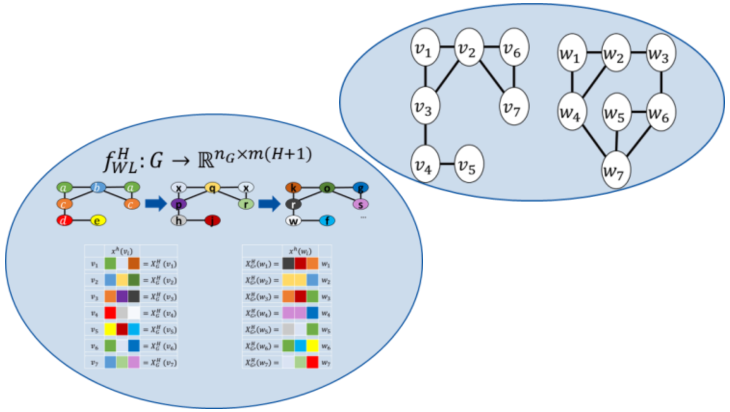
Solution: $\mathfrak{R}$-Convolution kernels

Problem: 1. Naive **aggregation** of **substructures** may disregard valuable information

2. Only a few approaches extendible for **continuously labelled** graphs



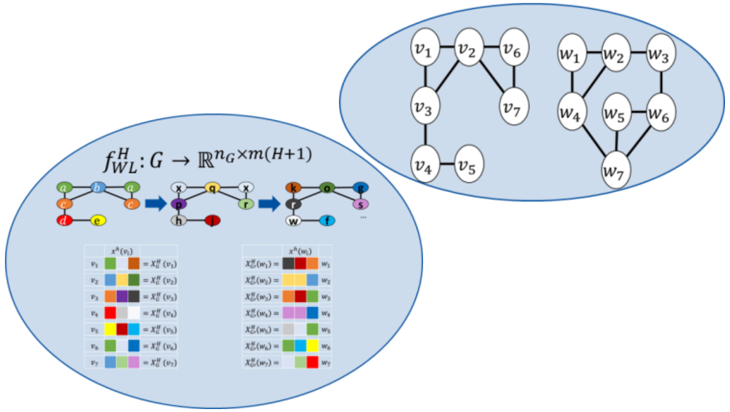
Method Overview



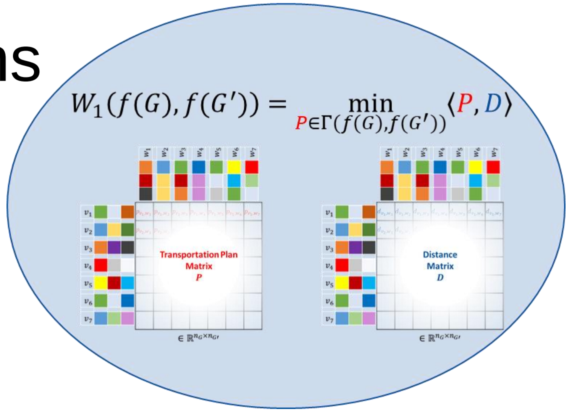
1. Transform each graph into a set of node embeddings
   Weisfeiler-Lehman node embedding scheme
   1. Idea / example
   2. Theory
   3. Example

Method Overview

1. Transform each graph into a set of node embeddings
   Weisfeiler-Lehman node embedding scheme
   1. Idea / example
   2. Theory
   3. Example

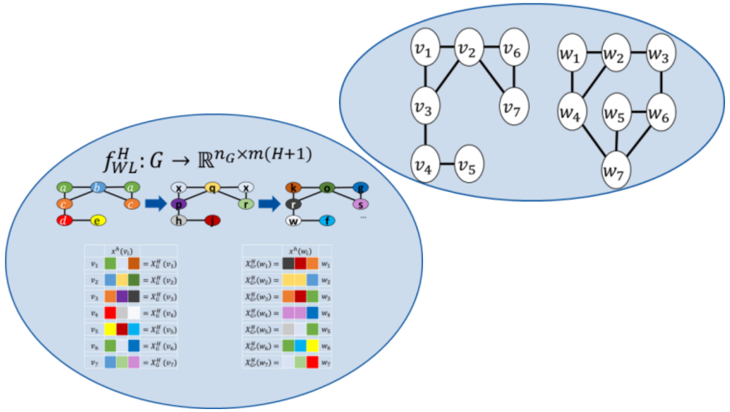


2. Measure the **Wasserstein distance** between each pair of graphs
   1. Transportation plan
   2. Distance measure
   3. Wasserstein distance

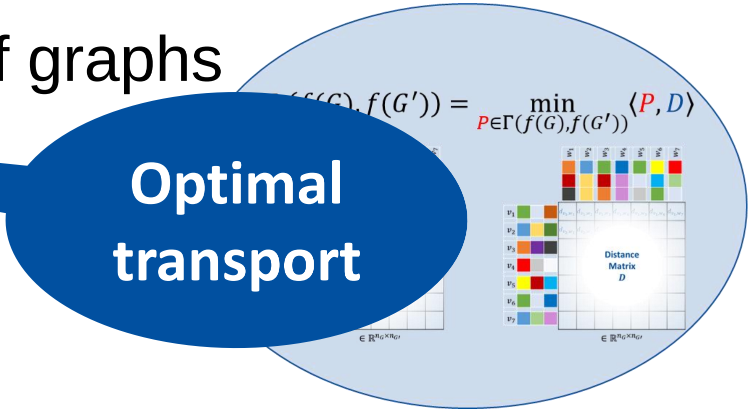


Method Overview

UNIVERSITÄT BONN

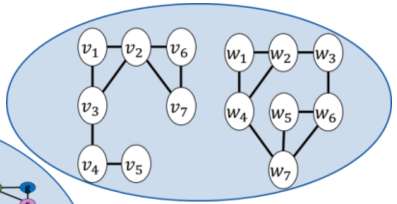


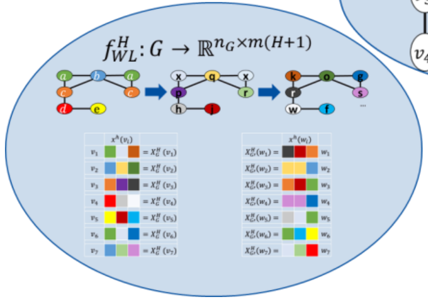
1. Transform each graph into a set of node embeddings
   Weisfeiler-Lehman node embedding scheme
   1. Idea / example
   2. Theory
   3. Example
2. Measure the **Wasserstein distance** between each pair of graphs
   1. Transportation plan
   2. Distance measure
   3. Wasserstein distance



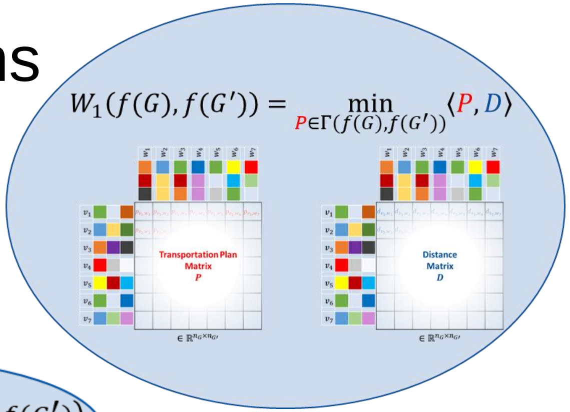
Method Overview

UNIVERSITÄT BONN





1. Transform each graph into a set of node embeddings
   Weisfeiler-Lehman node embedding scheme
    1. Idea / example
    2. Theory
    3. Example

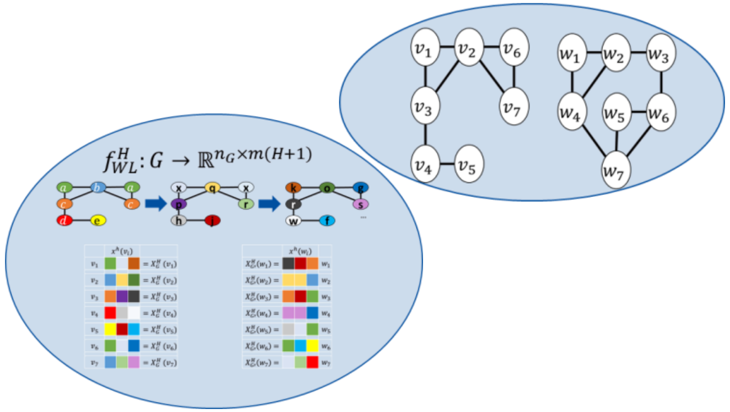


2. Measure the **Wasserstein distance** between each pair of graphs
    1. Transportation plan
    2. Distance measure
    3. Wasserstein distance
3. Wasserstein Weisfeiler-Lehman **kernel definition**

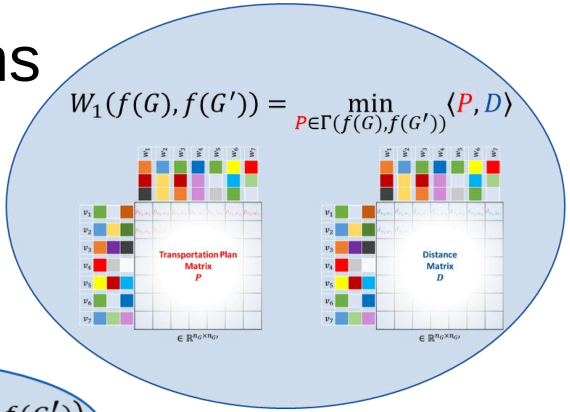
$$K_{WWL} = e^{-\lambda W_1(f(G), f(G'))}$$

Method Overview

UNIVERSITÄT BONN



1. Transform each graph into a set of node embeddings
   Weisfeiler-Lehman node embedding scheme
   1. Idea / example
   2. Theory
   3. Example

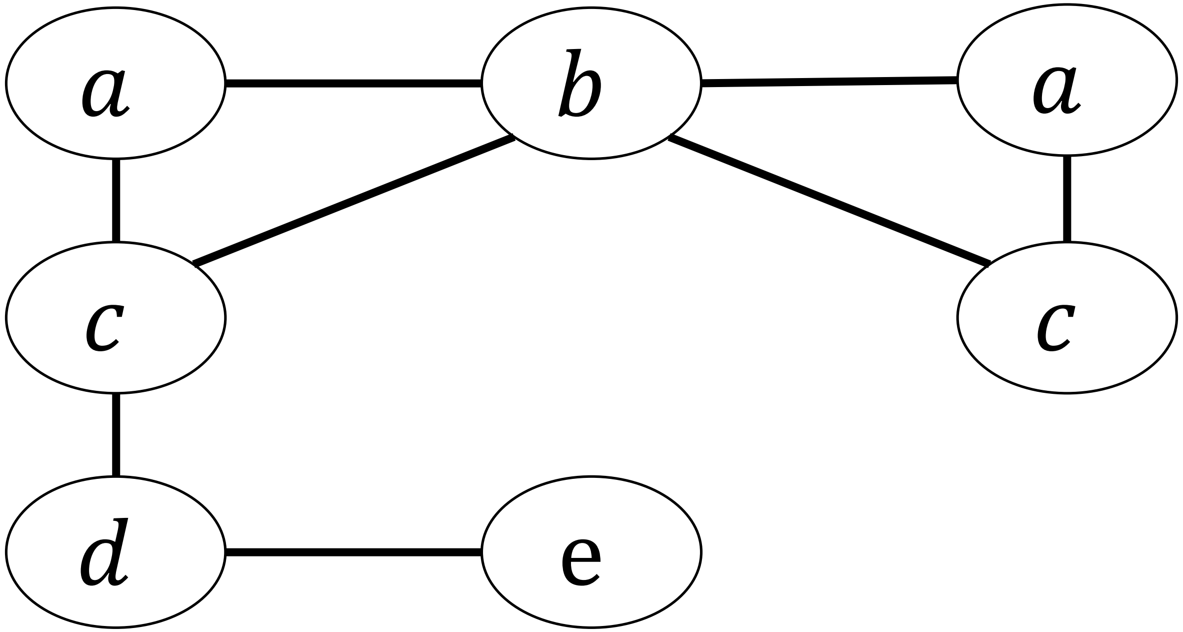


2. Measure the **Wasserstein distance** between each pair of graphs
   1. Transportation plan
   2. Distance measure
   3. Wasserstein distance
3. Wasserstein Weisfeiler-Lehman **kernel definition**

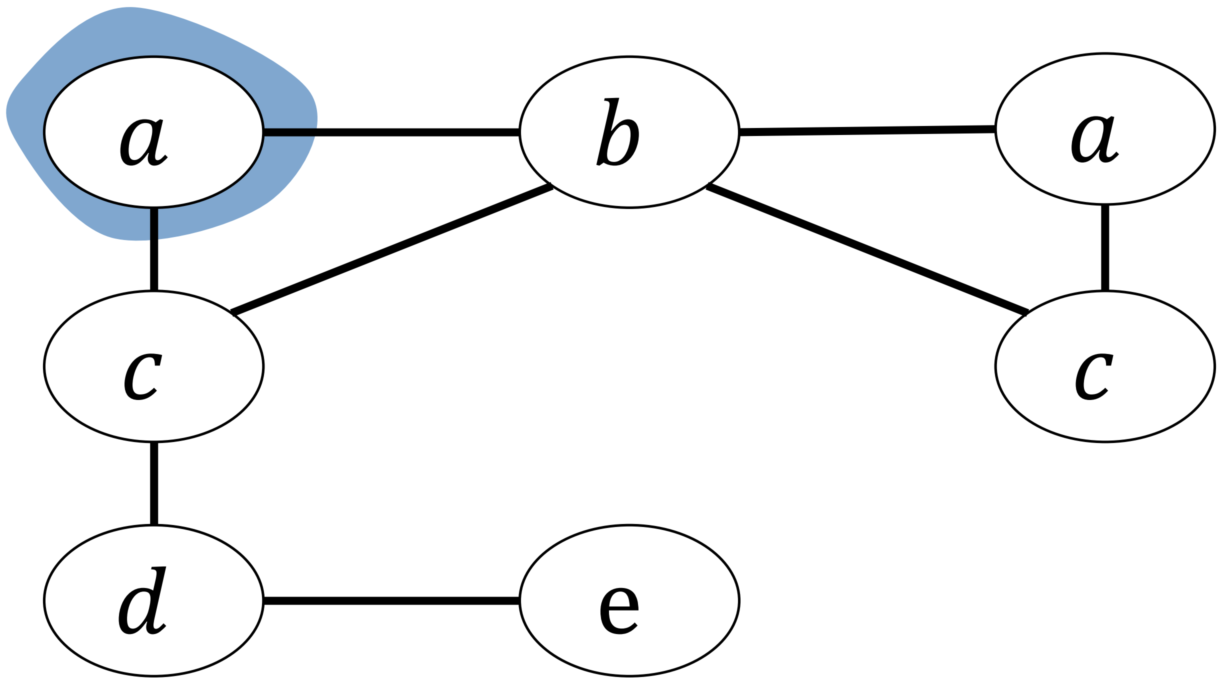
$$K_{WWL} = e^{-\lambda W_1(f(G), f(G'))}$$

(Differentiate between a finite set of **discrete** node labels and **continuously** valued node labels)

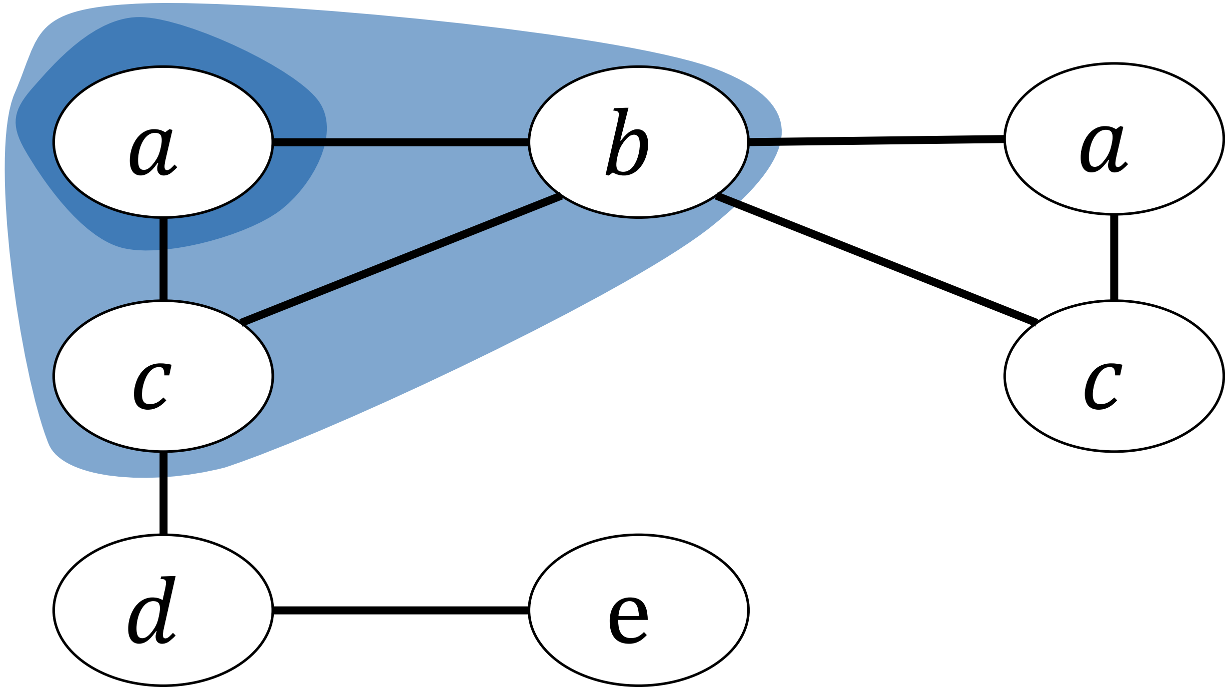
1.1 Weisfeiler-Lehman node embedding scheme
Idea



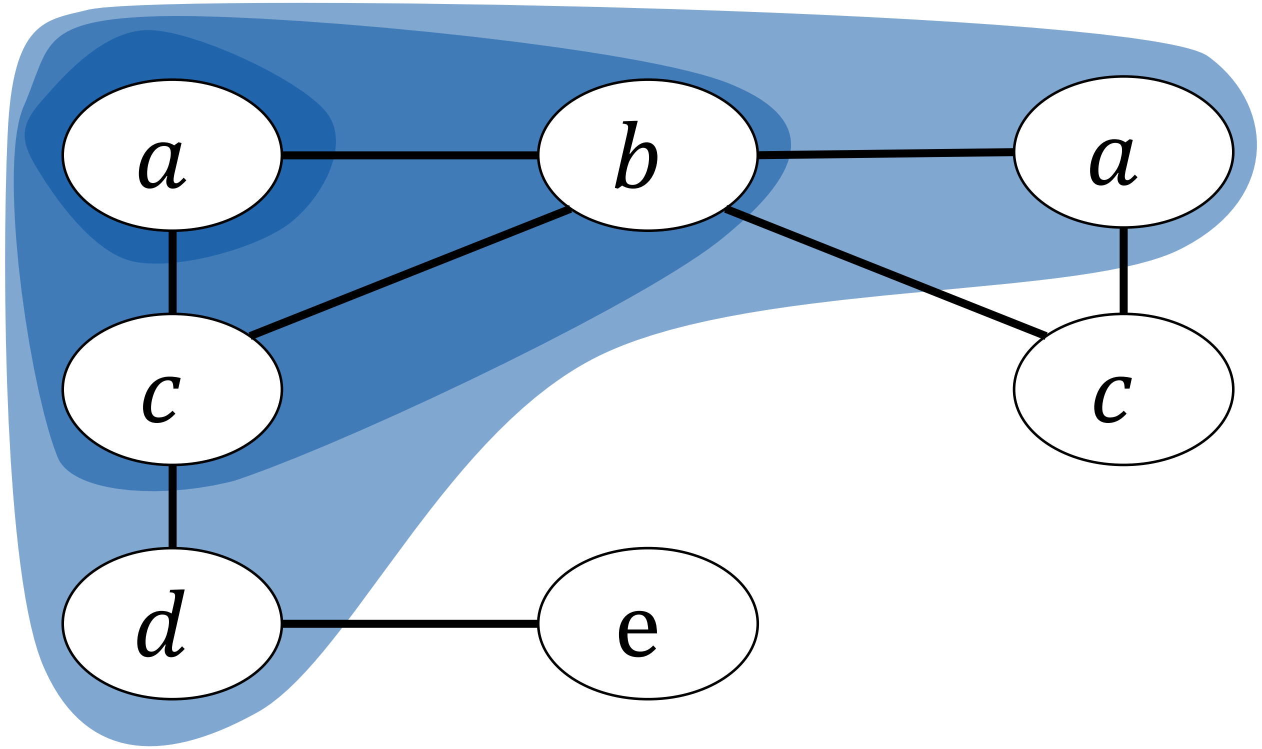
1.1 Weisfeiler-Lehman node embedding scheme
Idea



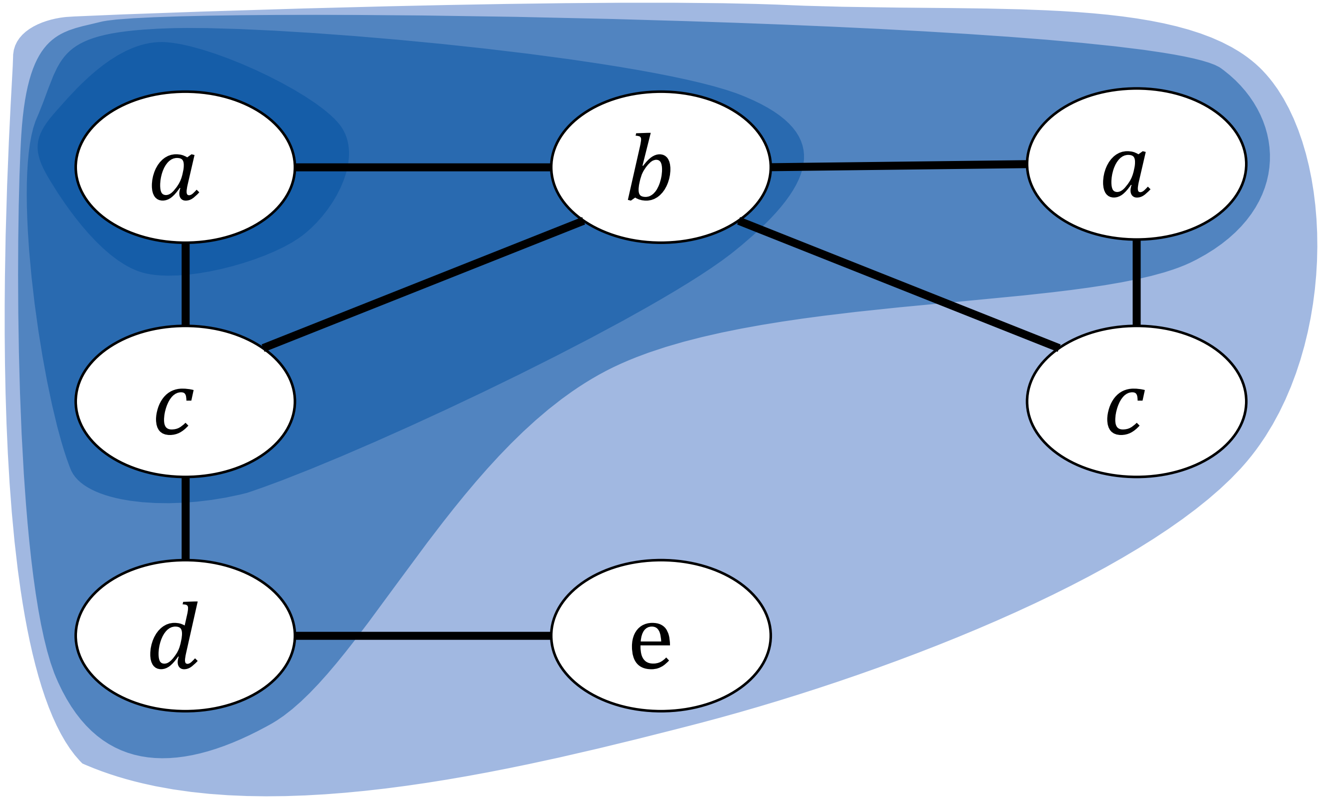
1.1 Weisfeiler-Lehman node embedding scheme
Idea



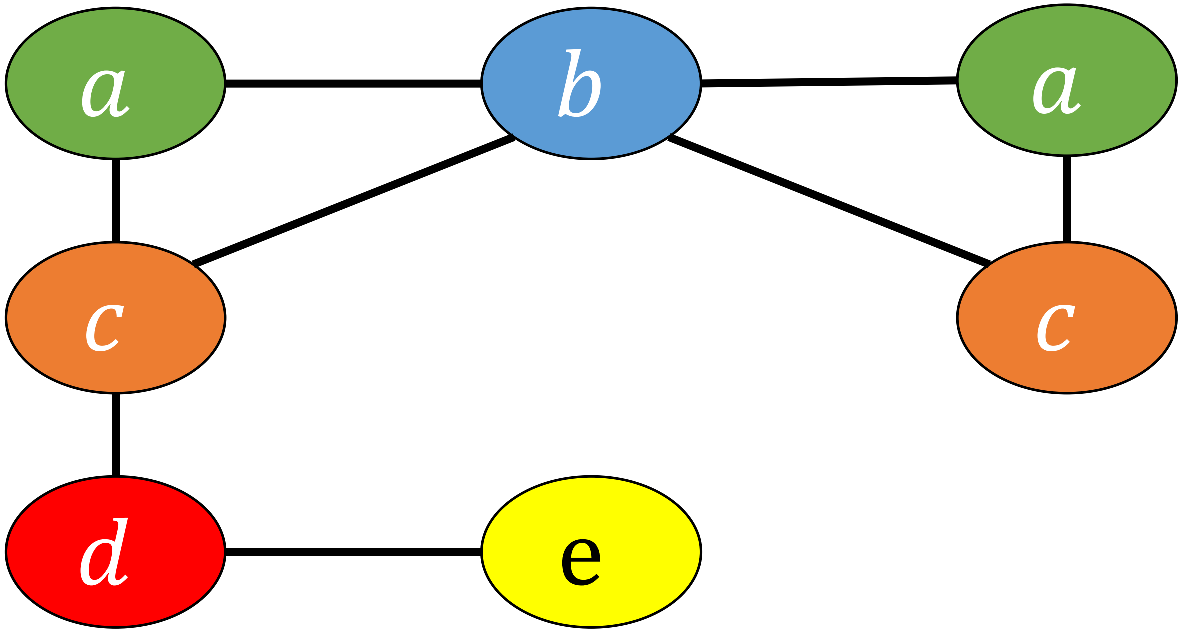
1.1 Weisfeiler-Lehman node embedding scheme
Idea



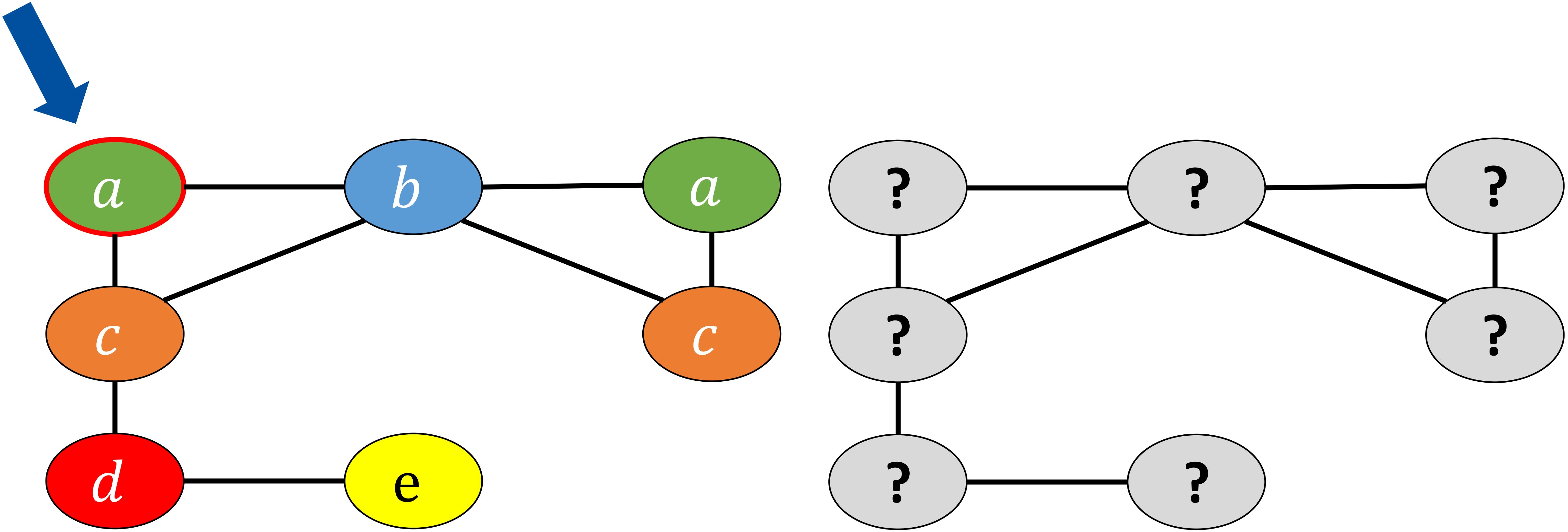
1.1 Weisfeiler-Lehman node embedding scheme
Idea



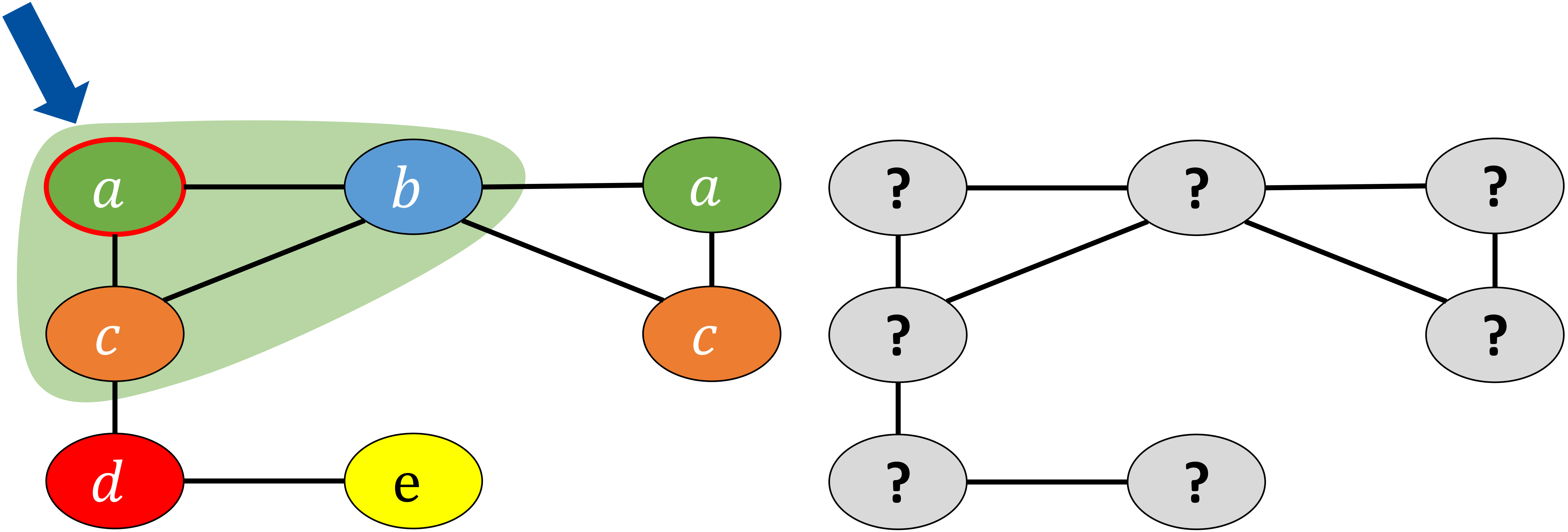
1.1 Weisfeiler-Lehman node embedding scheme
Idea / Example



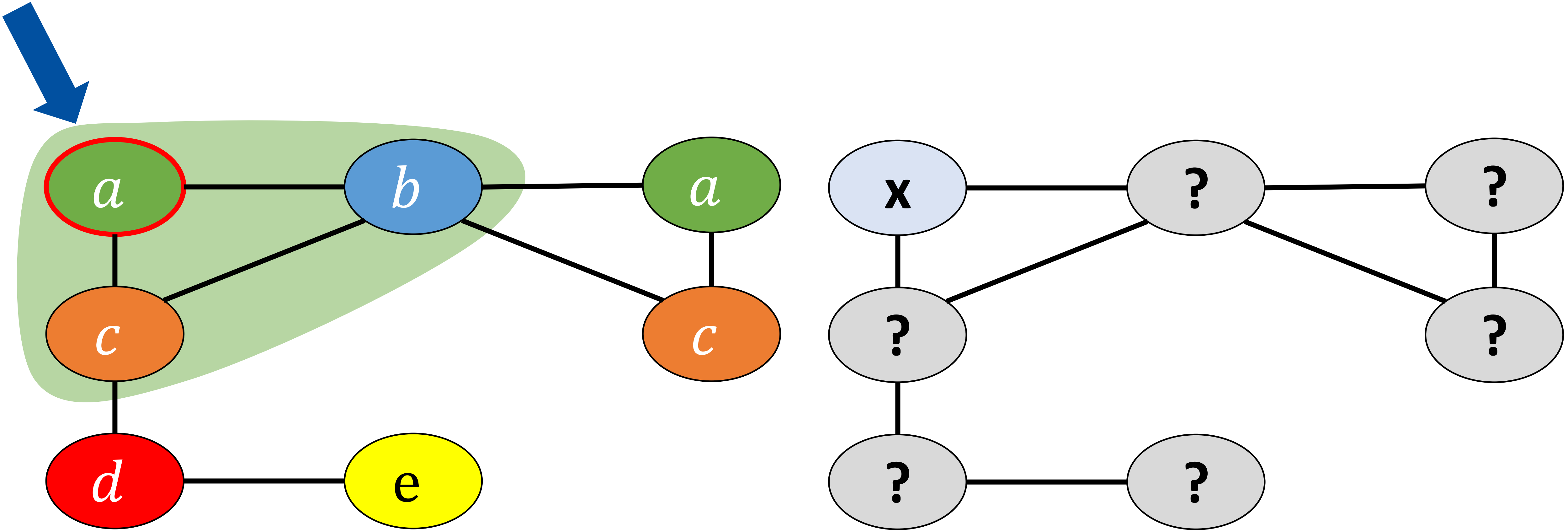
1.1 Weisfeiler-Lehman node embedding scheme
Idea / Example



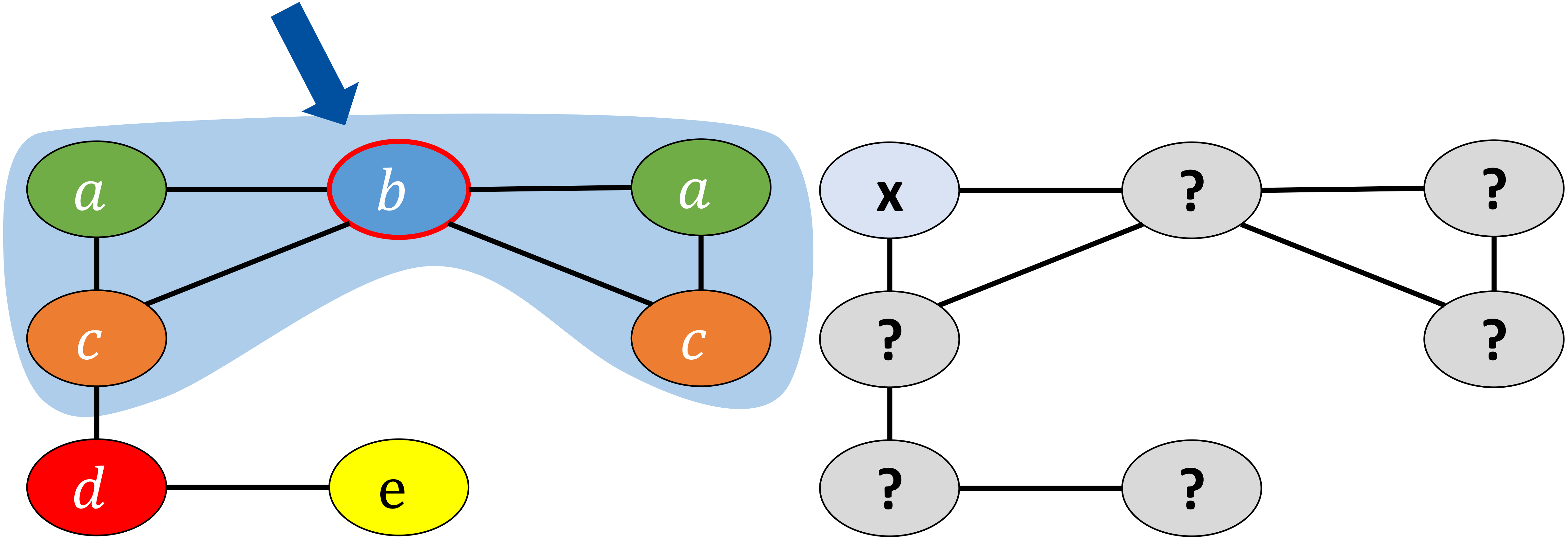
1.1 Weisfeiler-Lehman node embedding scheme
Idea / Example



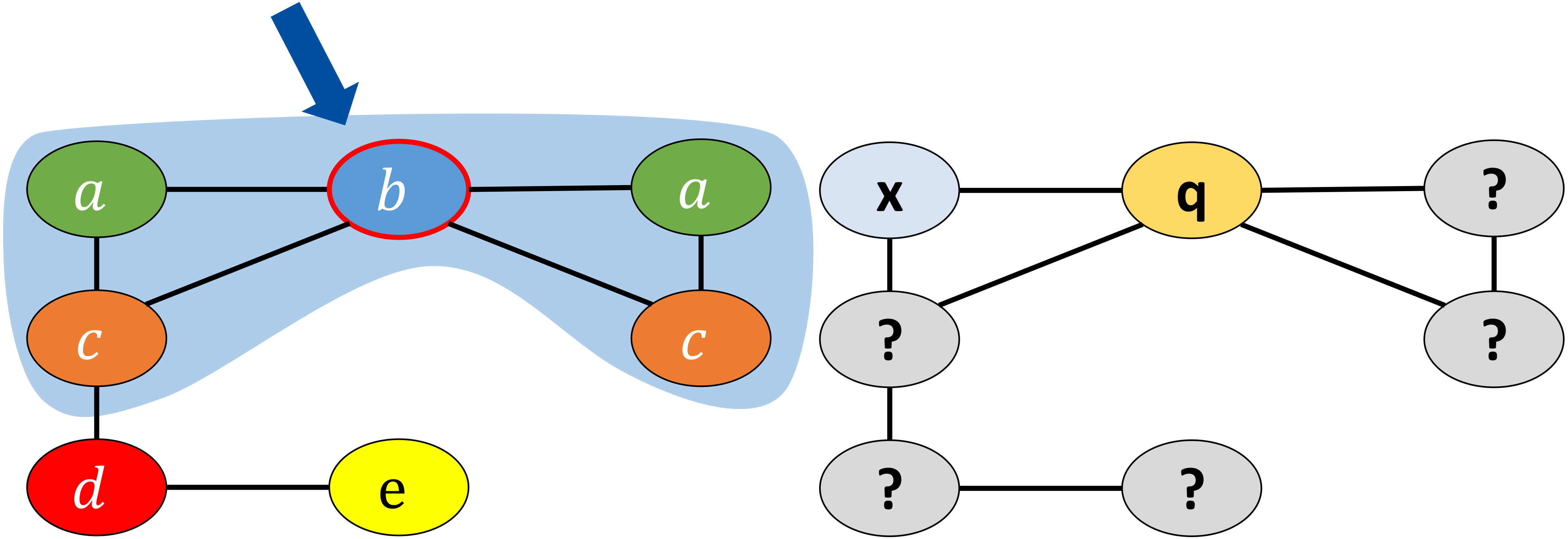
1.1 Weisfeiler-Lehman node embedding scheme
Idea / Example



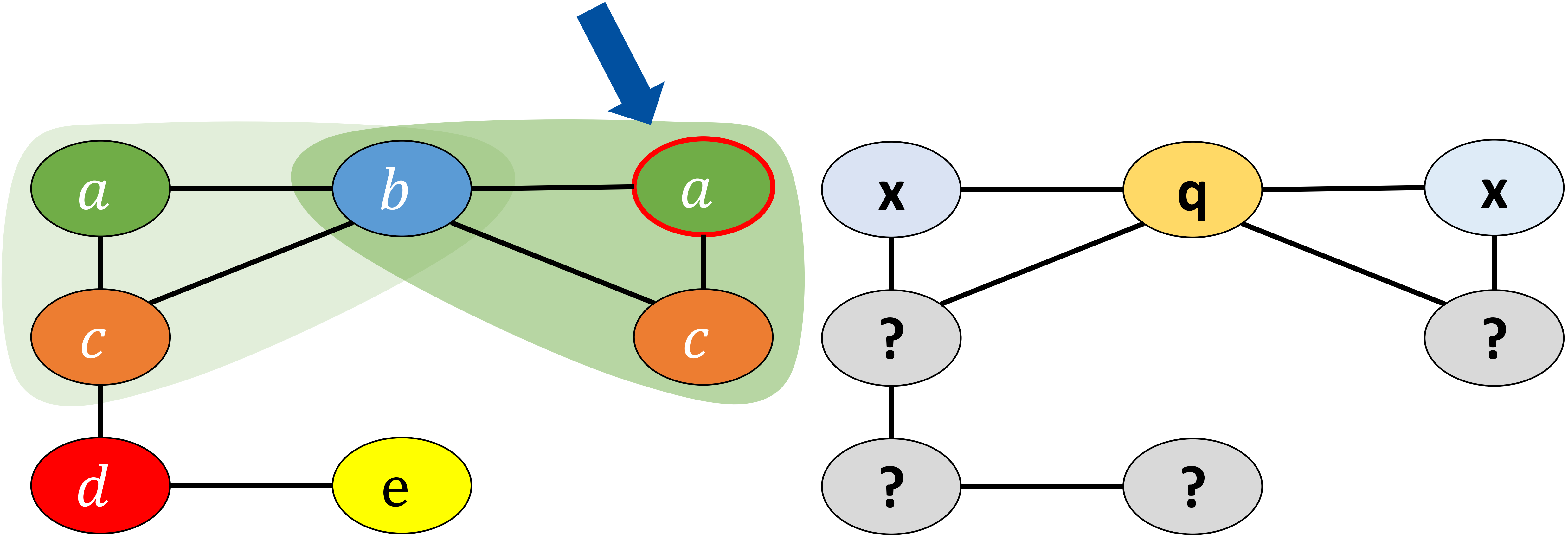
1.1 Weisfeiler-Lehman node embedding scheme
Idea / Example



1.1 Weisfeiler-Lehman node embedding scheme
Idea / Example



1.1 Weisfeiler-Lehman node embedding scheme
Idea / Example



1.1 Weisfeiler-Lehman node embedding scheme Definition

m: dimensionality of the node labels

n_G: cardinality of the node set

$N(v)$: neighbourhood of v

DISCRETE node labels	CONTINUOUS node labels
$x^{h+1}(v) = hash(x^h(v), N^h(v))$	

1.1 Weisfeiler-Lehman node embedding scheme Definition

m: dimensionality of the node labels

n_G: cardinality of the node set

$N(v)$: neighbourhood of v

$$N^h(v) = \{x^h(u_1), \dots, x^h(u_{|N(v)|})\}$$

DISCRETE node labels	CONTINUOUS node labels
$x^{h+1}(v) = hash(x^h(v), N^h(v))$	

1.1 Weisfeiler-Lehman node embedding scheme Definition

UNIVERSITÄT BONN

m: dimensionality of the node labels

n_G: cardinality of the node set

$N(v)$: neighbourhood of v

$$N^h(v) = \{x^h(u_1), \dots, x^h(u_{|N(v)|})\}$$

DISCRETE node labels	CONTINUOUS node labels
$x^{h+1}(v) = hash(x^h(v), N^h(v))$	$x^{h+1}(v) = \frac{1}{2}\left(x^h(v) + \frac{1}{\deg(v)} \sum_{u \in N(v)} w((u,v))\, x^h(u)\right)$

$(h+1)$-Neighbourhood embedding of v

1.1 Weisfeiler-Lehman node embedding scheme
Definition

m: dimensionality of the node labels

n_G: cardinality of the node set

$N(v)$: neighbourhood of v

$$N^h(v) = \{x^h(u_1), \dots, x^h(u_{|N(v)|})\}$$

DISCRETE node labels	CONTINUOUS node labels
$x^{h+1}(v) = hash(x^h(v), N^h(v))$	$x^{h+1}(v) = \frac{1}{2}\left(x^h(v) + \frac{1}{\deg(v)} \sum_{u \in N(v)} w((u,v))\, x^h(u)\right)$ $\in \mathbb{R}^m$

Node embedding of v_1

$$X_G^H(v_1) = [x^0(v_1), \dots, x^H(v_1)] \quad \in \mathbb{R}^{m(H+1)}$$

1.1 Weisfeiler-Lehman node embedding scheme Definition

UNIVERSITÄT BONN

m: dimensionality of the node labels

n_G: cardinality of the node set

$N(v)$: neighbourhood of v

$$N^h(v) = \{x^h(u_1), \dots, x^h(u_{|N(v)|})\}$$

DISCRETE node labels	CONTINUOUS node labels
$x^{h+1}(v) = hash(x^h(v), N^h(v))$	$x^{h+1}(v) = \frac{1}{2}\left(x^h(v) + \frac{1}{\deg(v)} \sum_{u \in N(v)} w((u,v))\, x^h(u)\right) \in \mathbb{R}^m$

Node embedding of v_1 $\quad X_G^H(v_1) = [x^0(v_1), \dots, x^H(v_1)] \quad \in \mathbb{R}^{m(H+1)}$

Graph embedding $\quad f^H: G \longrightarrow \left(X_G^H(v_1), \dots, X_G^H(v_{n_G})\right)^T \quad \in \mathbb{R}^{n_G \times m(H+1)}$

1.1 Weisfeiler-Lehman node embedding scheme
Example – Node labels

$m = 1$: dimensionality of the node labels

$n_G = 7$: cardinality of the node set

$N(v)$: neighbourhood of v

$$x^{h+1}(v) = hash(x^h(v), N^h(v))$$

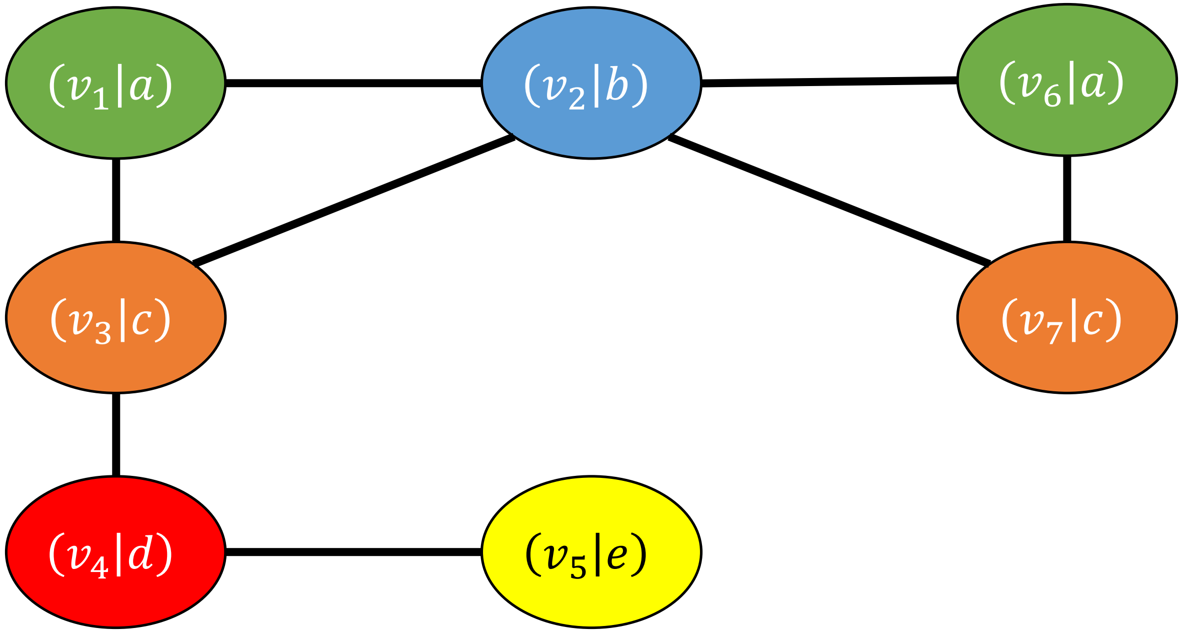
1.1 Weisfeiler-Lehman node embedding scheme
Example – Node labels

$m = 1$: dimensionality of the node labels

$n_G = 7$: cardinality of the node set

$N(v)$: neighbourhood of v

$$x^{h+1}(v) = hash(x^h(v), N^h(v))$$



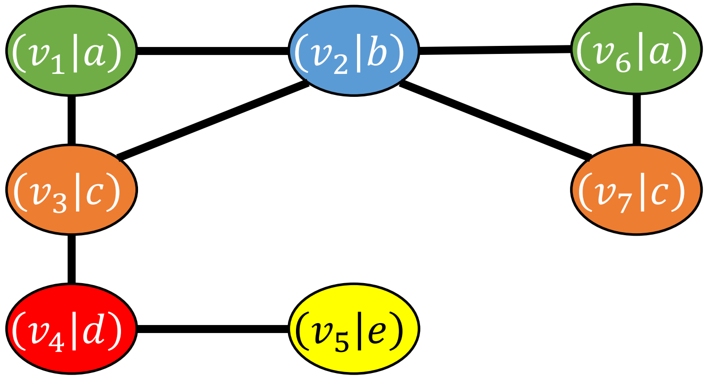
1.1 Weisfeiler-Lehman node embedding scheme Example – Node labels

$m = 1$: dimensionality of the node labels

$n_G = 7$: cardinality of the node set

$N(v)$: neighbourhood of v

$$x^{h+1}(v) = hash(x^h(v), N^h(v))$$



i	$x^0(v_i)$	$x^1(v_i)$	$x^2(v_i)$	…	$x^H(v_i)$
1	a			…	…
2	b			…	…
3	c			…	…
4	d			…	…
5	e			…	…
6	a			…	…
7	c			…	…

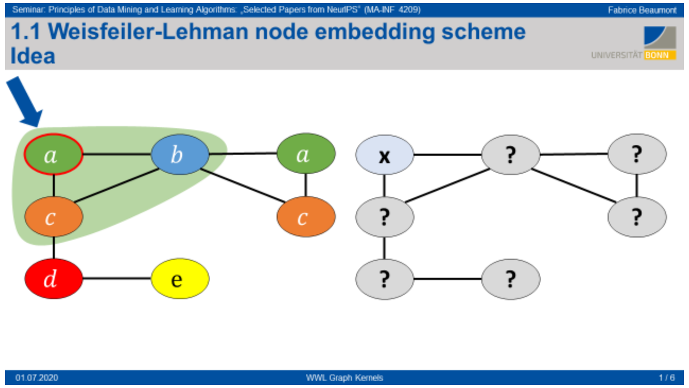
1.1 Weisfeiler-Lehman node embedding scheme
Example – Node labels

$m = 1$: dimensionality of the node labels

$n_G = 7$: cardinality of the node set

$N(v)$: neighbourhood of v

$$x^{h+1}(v) = hash(x^h(v), N^h(v))$$



i	$x^0(v_i)$	$x^1(v_i)$	$x^2(v_i)$	…	$x^H(v_i)$
1	a	$\text{hash}(a, \{b, c\}) = x$		…	…
2	b			…	…
3	c			…	…
4	d			…	…
5	e			…	…
6	a			…	…
7	c			…	…

1.1 Weisfeiler-Lehman node embedding scheme
Example – Node labels

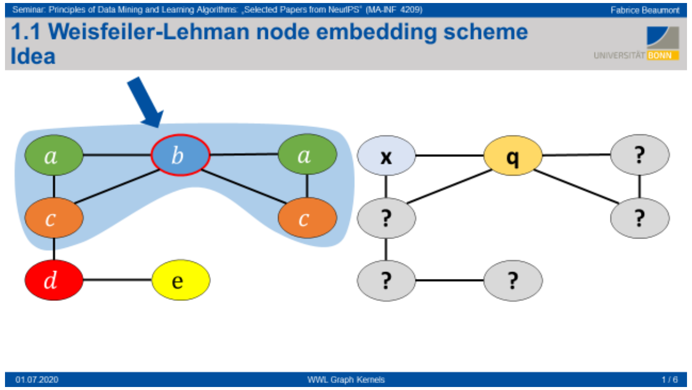
UNIVERSITÄT BONN

$m = 1$: dimensionality of the node labels

$n_G = 7$: cardinality of the node set

$N(v)$: neighbourhood of v

$$x^{h+1}(v) = hash(x^h(v), N^h(v))$$



i	$x^0(v_i)$	$x^1(v_i)$	$x^2(v_i)$	…	$x^H(v_i)$
1	a	hash$(a, \{b, c\}) = x$		…	…
2	b	hash$(b, \{a, c\}) = q$		…	…
3	c			…	…
4	d			…	…
5	e			…	…
6	a			…	…
7	c			…	…

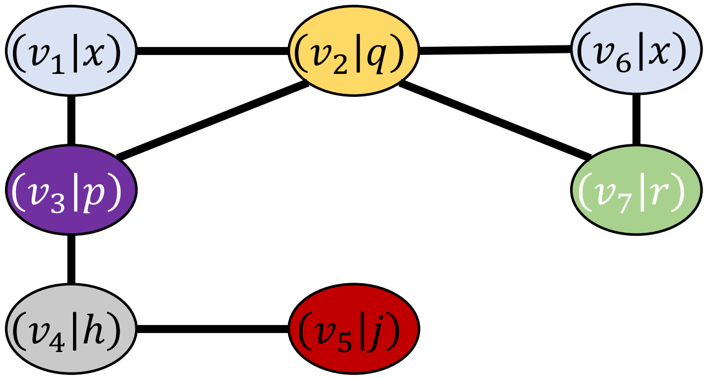
1.1 Weisfeiler-Lehman node embedding scheme
Example – Node labels

$m = 1$: dimensionality of the node labels

$n_G = 7$: cardinality of the node set

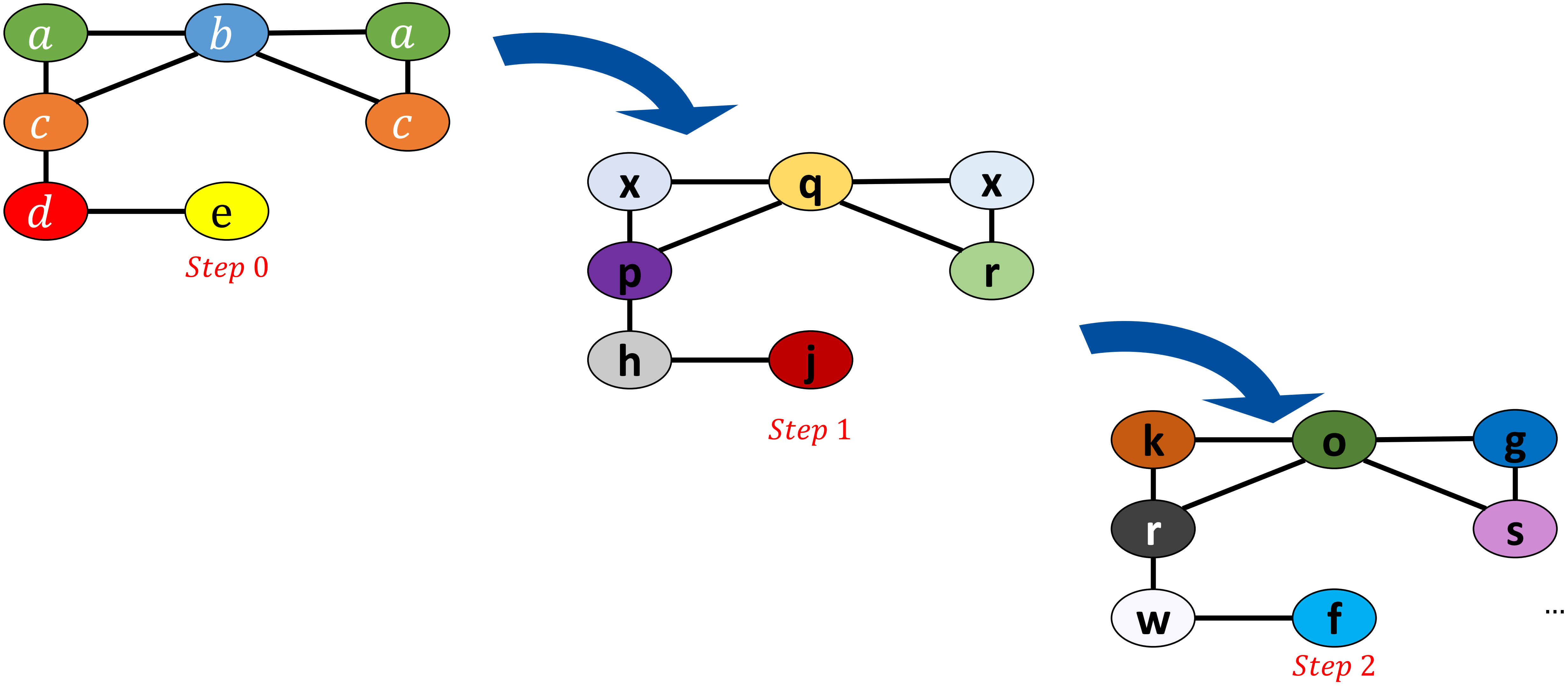
$N(v)$: neighbourhood of v

$$x^{h+1}(v) = hash(x^h(v), N^h(v))$$



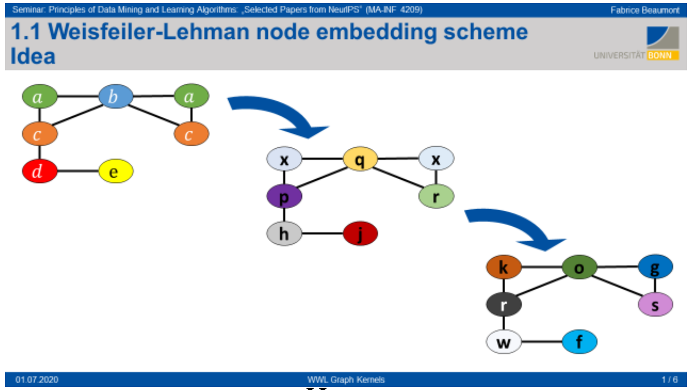
i	$x^0(v_i)$	$x^1(v_i)$	$x^2(v_i)$	…	$x^H(v_i)$
1	a	hash$(a, \{b, c\}) = x$		…	…
2	b	hash$(b, \{a, c\}) = q$		…	…
3	c	hash$(c, \{a, b, d\}) = p$		…	…
4	d	hash$(d, \{c, e\}) = h$		…	…
5	e	hash$(e, \{d\}) = j$		…	…
6	a	hash$(a, \{b, c\}) = x$		…	…
7	c	hash$(c, \{a, b\}) = r$		…	…

1.1 Weisfeiler-Lehman node embedding scheme
Example – Node labels



1.1 Weisfeiler-Lehman node embedding scheme
Example – Node labels

UNIVERSITÄT BONN



$m = 1$: dimensionality of the node labels

$n_G = 7$: cardinality of the node set

$N(v)$: neighbourhood of v

$$x^{h+1}(v) = hash(x^h(v), N^h(v))$$

i	$x^0(v_i)$	$x^1(v_i)$	$x^2(v_i)$	…	$x^H(v_i)$
1	a	hash$(a, \{b, c\}) = x$	hash$(x, \{q, p\}) = k$	…	…
2	b	hash$(b, \{a, c\}) = q$	hash$(q, \{x, p\}) = o$	…	…
3	c	hash$(c, \{a, b, d\}) = p$	hash$(p, \{x, q, h\}) = r$	…	…
4	d	hash$(d, \{c, e\}) = h$	hash$(h, \{p, j\}) = w$	…	…
5	e	hash$(e, \{d\}) = j$	hash$(j, \{h\}) = f$	…	…
6	a	hash$(a, \{b, c\}) = x$	hash$(x, \{q, r\}) = g$	…	…
7	c	hash$(c, \{a, b\}) = r$	hash$(r, \{x, q\}) = s$	…	…

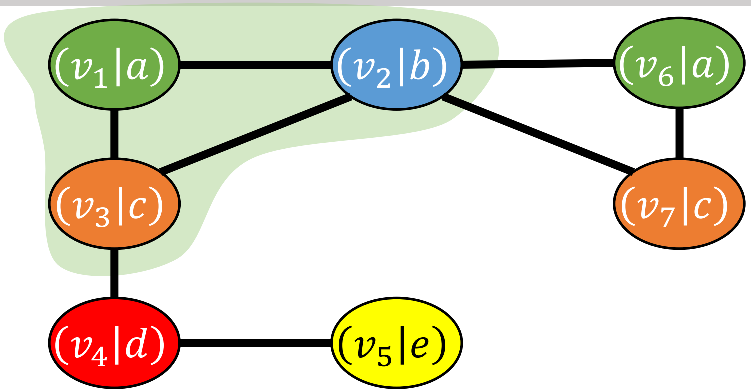
1.1 Weisfeiler-Lehman node embedding scheme
Example – Node labels

$m = 1$: dimensionality of the node labels

$n_G = 7$: cardinality of the node set

$N(v)$: neighbourhood of v

$$x^{h+1}(v) = hash(x^h(v), N^h(v))$$



i	$x^0(v_i)$	$x^1(v_i)$	$x^2(v_i)$	...	$x^H(v_i)$
1	a	$\text{hash}(a, \{b, c\}) = x$	$\text{hash}(x, \{q, p\}) = k$	...	...
2	b	$\text{hash}(b, \{a, c\}) = q$	$\text{hash}(q, \{x, p\}) = o$	...	...
3	c	$\text{hash}(c, \{a, b, d\}) = p$	$\text{hash}(p, \{x, q, h\}) = r$	...	...
4	d	$\text{hash}(d, \{c, e\}) = h$	$\text{hash}(h, \{p, j\}) = w$	...	...
5	e	$\text{hash}(e, \{d\}) = j$	$\text{hash}(j, \{h\}) = f$	...	...
6	a	$\text{hash}(a, \{b, c\}) = x$	$\text{hash}(x, \{q, r\}) = g$	...	...
7	c	$\text{hash}(c, \{a, b\}) = r$	$\text{hash}(r, \{x, q\}) = s$	...	...

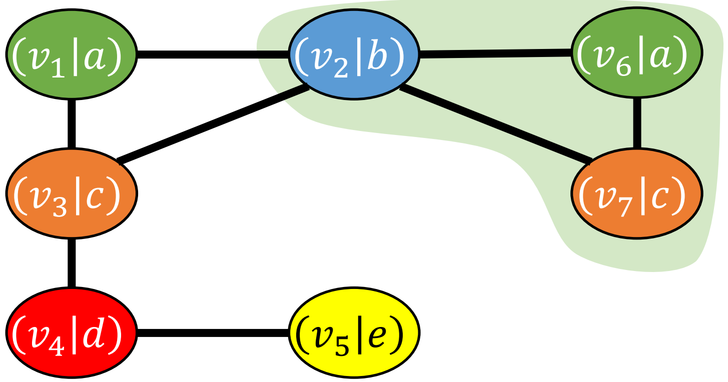
1.1 Weisfeiler-Lehman node embedding scheme Example – Node labels

$m = 1$: dimensionality of the node labels

$n_G = 7$: cardinality of the node set

$N(v)$: neighbourhood of v

$$x^{h+1}(v) = hash(x^h(v), N^h(v))$$



i	$x^0(v_i)$	$x^1(v_i)$	$x^2(v_i)$	…	$x^H(v_i)$
1	a	$\text{hash}(a, \{b, c\}) = x$	$\text{hash}(x, \{q, p\}) = k$	…	…
2	b	$\text{hash}(b, \{a, c\}) = q$	$\text{hash}(q, \{x, p\}) = o$	…	…
3	c	$\text{hash}(c, \{a, b, d\}) = p$	$\text{hash}(p, \{x, q, h\}) = r$	…	…
4	d	$\text{hash}(d, \{c, e\}) = h$	$\text{hash}(h, \{p, j\}) = w$	…	…
5	e	$\text{hash}(e, \{d\}) = j$	$\text{hash}(j, \{h\}) = f$	…	…
6	a	$\text{hash}(a, \{b, c\}) = x$	$\text{hash}(x, \{q, r\}) = g$	…	…
7	c	$\text{hash}(c, \{a, b\}) = r$	$\text{hash}(r, \{x, q\}) = s$	…	…

1.1 Weisfeiler-Lehman node embedding scheme Example – Node labels

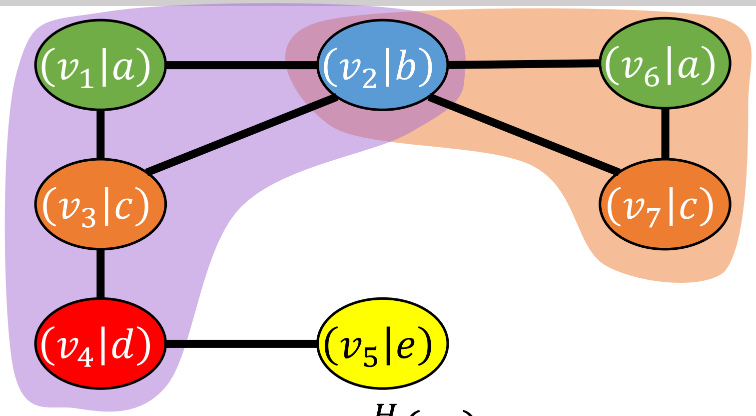
UNIVERSITÄT BONN

$m = 1$: dimensionality of the node labels

$n_G = 7$: cardinality of the node set

$N(v)$: neighbourhood of v

$$x^{h+1}(v) = hash(x^h(v), N^h(v))$$



i	$x^0(v_i)$	$x^1(v_i)$	$x^2(v_i)$	…	$x^H(v_i)$
1	a	hash$(a,\{b,c\}) = x$	hash$(x,\{q,p\}) = k$	…	…
2	b	hash$(b,\{a,c\}) = q$	hash$(q,\{x,p\}) = o$	…	…
3	c	hash$(c,\{a,b,d\}) = p$	hash$(p,\{x,q,h\}) = r$	…	…
4	d	hash$(d,\{c,e\}) = h$	hash$(h,\{p,j\}) = w$	…	…
5	e	hash$(e,\{d\}) = j$	hash$(j,\{h\}) = f$	…	…
6	a	hash$(a,\{b,c\}) = x$	hash$(x,\{q,r\}) = g$	…	…
7	c	hash$(c,\{a,b\}) = r$	hash$(r,\{x,q\}) = s$	…	…

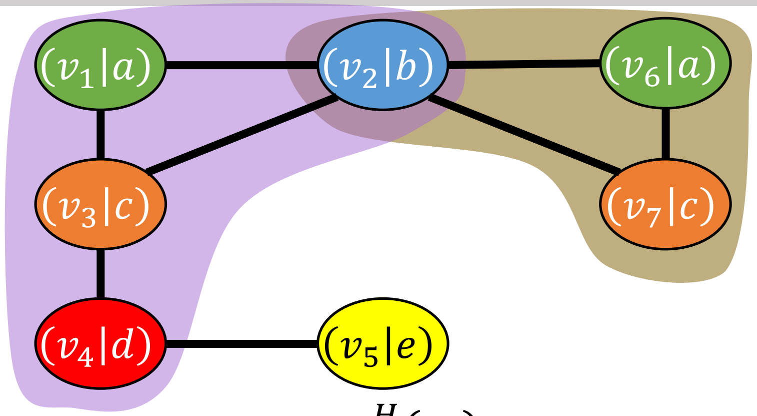
1.1 Weisfeiler-Lehman node embedding scheme
Example – Node labels

$m = 1$: dimensionality of the node labels

$n_G = 7$: cardinality of the node set

$N(v)$: neighbourhood of v

$$x^{h+1}(v) = hash(x^h(v), N^h(v))$$



i	$x^0(v_i)$	$x^1(v_i)$	$x^2(v_i)$	…	$x^H(v_i)$
1	a	hash$(a, \{b, c\}) = x$	hash$(x, \{q, \mathbf{p}\}) = \boldsymbol{k}$	…	…
2	b	hash$(b, \{a, c\}) = q$	hash$(q, \{x, p\}) = o$	…	…
3	c	hash$(c, \{a, b, d\}) = p$	hash$(p, \{x, q, h\}) = r$	…	…
4	d	hash$(d, \{c, e\}) = h$	hash$(h, \{p, j\}) = w$	…	…
5	e	hash$(e, \{d\}) = j$	hash$(j, \{h\}) = f$	…	…
6	a	hash$(a, \{b, c\}) = x$	hash$(x, \{q, \mathbf{r}\}) = \boldsymbol{g}$	…	…
7	c	hash$(c, \{a, b\}) = r$	hash$(r, \{x, q\}) = s$	…	…

1.1 Weisfeiler-Lehman node embedding scheme Example – Node labels

$m = 1$: dimensionality of the node labels

$n_G = 7$: cardinality of the node set

$N(v)$: neighbourhood of v

$$x^{h+1}(v) = hash(x^h(v), N^h(v))$$

i	$x^0(v_i)$	$x^1(v_i)$	$x^2(v_i)$	…	$x^H(v_i)$	
1	a	$\text{hash}(a, \{b, c\}) = x$	$\text{hash}(x, \{q, p\}) = k$	…	…	$X_G^H(v_1)$
2	b	$\text{hash}(b, \{a, c\}) = q$	$\text{hash}(q, \{x, p\}) = o$	…	…	$X_G^H(v_2)$
3	c	$\text{hash}(c, \{a, b, d\}) = p$	$\text{hash}(p, \{x, q, h\}) = r$	…	…	$X_G^H(v_3)$
4	d	$\text{hash}(d, \{c, e\}) = h$	$\text{hash}(h, \{p, j\}) = w$	…	…	$X_G^H(v_4)$
5	e	$\text{hash}(e, \{d\}) = j$	$\text{hash}(j, \{h\}) = f$	…	…	$X_G^H(v_5)$
6	a	$\text{hash}(a, \{b, c\}) = x$	$\text{hash}(x, \{q, r\}) = g$	…	…	$X_G^H(v_6)$
7	c	$\text{hash}(c, \{a, b\}) = r$	$\text{hash}(r, \{x, q\}) = s$	…	…	$X_G^H(v_7)$

1.1 Weisfeiler-Lehman node embedding scheme Example – Node labels

$m = 1$: dimensionality of the node labels

$n_G = 7$: cardinality of the node set

$N(v)$: neighbourhood of v

$$x^{h+1}(v) = hash(x^h(v), N^h(v))$$

Graph embedding $\boldsymbol{f^H}(G)$

2^{nd}-Neighbourhood embedding of v

i	$x^0(v_i)$	$x^1(v_i)$	$x^2(v_i)$	…	$x^H(v_i)$	
1	a	hash$(a, \{b, c\}) = x$	hash$(x, \{q, p\}) = k$	…	…	$X_G^H(v_1)$
2	b	hash$(b, \{a, c\}) = q$	hash$(q, \{x, p\}) = o$	…	…	$X_G^H(v_2)$
3	c	hash$(c, \{a, b, d\}) = p$	hash$(p, \{x, q, h\}) = r$	…	…	$X_G^H(v_3)$
4	d	hash$(d, \{c, e\}) = h$	hash$(h, \{p, j\}) = w$	…	…	$X_G^H(v_4)$
5	e	hash$(e, \{d\}) = j$	hash$(j, \{h\}) = f$	…	…	$X_G^H(v_5)$
6	a	hash$(a, \{b, c\}) = x$	hash$(x, \{q, r\}) = g$	…	…	$X_G^H(v_6)$
7	c	hash$(c, \{a, b\}) = r$	hash$(r, \{x, q\}) = s$	…	…	$X_G^H(v_7)$

Node embedding of v_7

Method Overview

UNIVERSITÄT BONN

1. Transform each graph into a set of node embeddings

   Weisfeiler-Lehman node embedding scheme

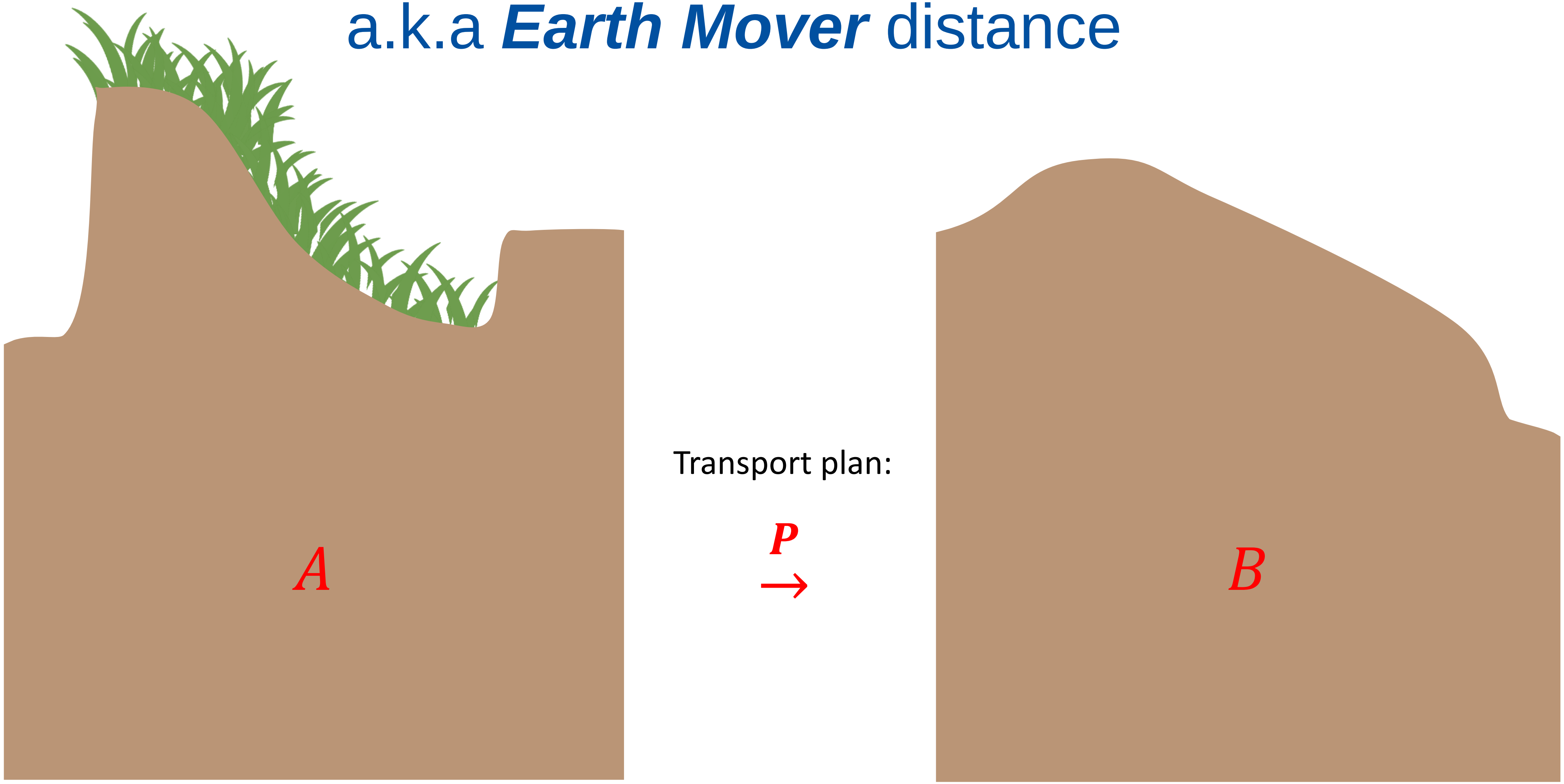
   1. Idea / example
   2. Theory
   3. Example

Method Overview

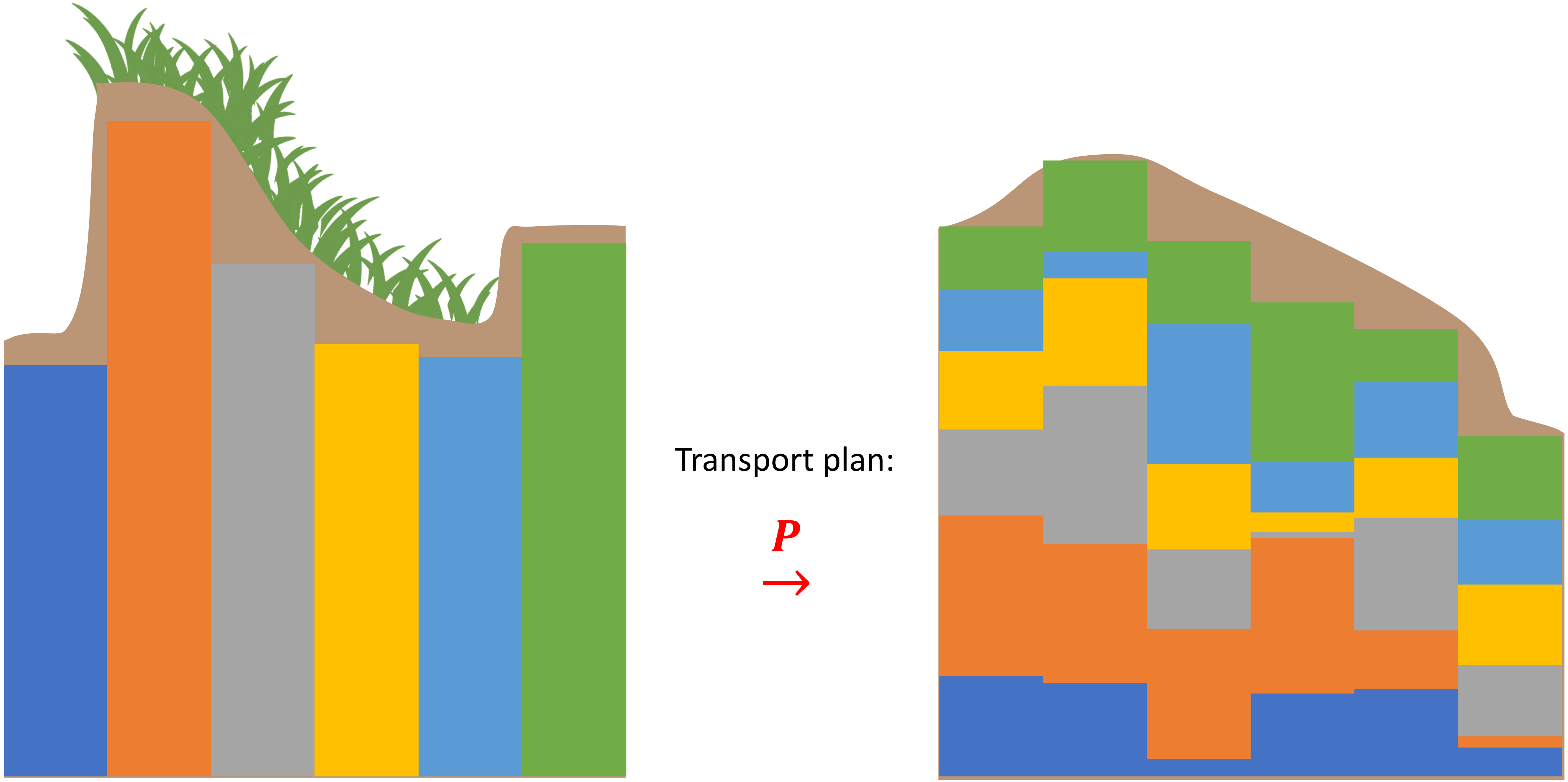
1. Transform each graph into a set of node embeddings
   Weisfeiler-Lehman node embedding scheme
    1. Idea / example
    2. Theory
    3. Example
2. Measure the **Wasserstein distance** between each pair of graphs
    1. Transportation plan
    2. Distance measure
    3. Wasserstein distance

2.2 Graph Wasserstein distance
Motivation

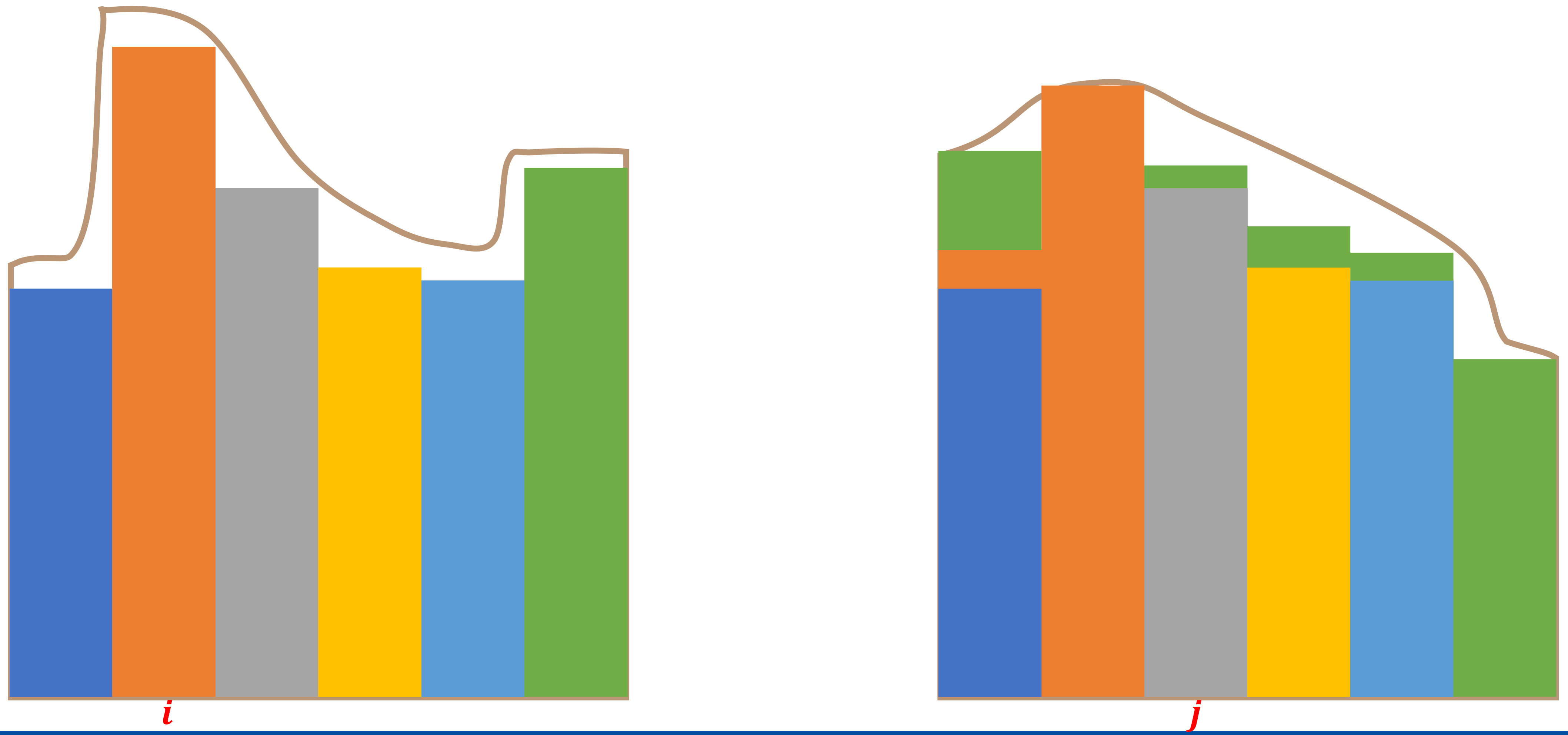
a.k.a ***Earth Mover*** distance



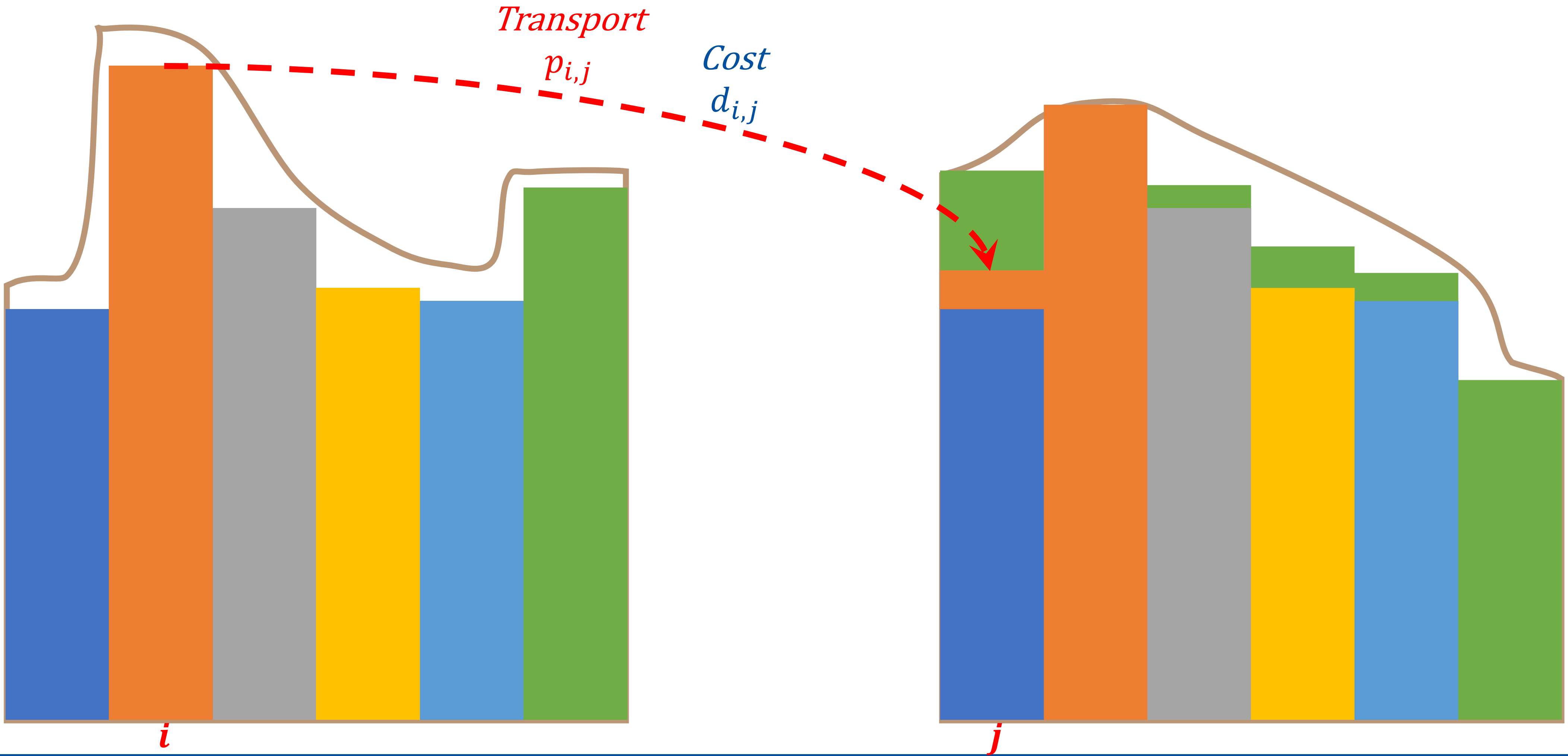
2.2 Graph Wasserstein distance
Motivation



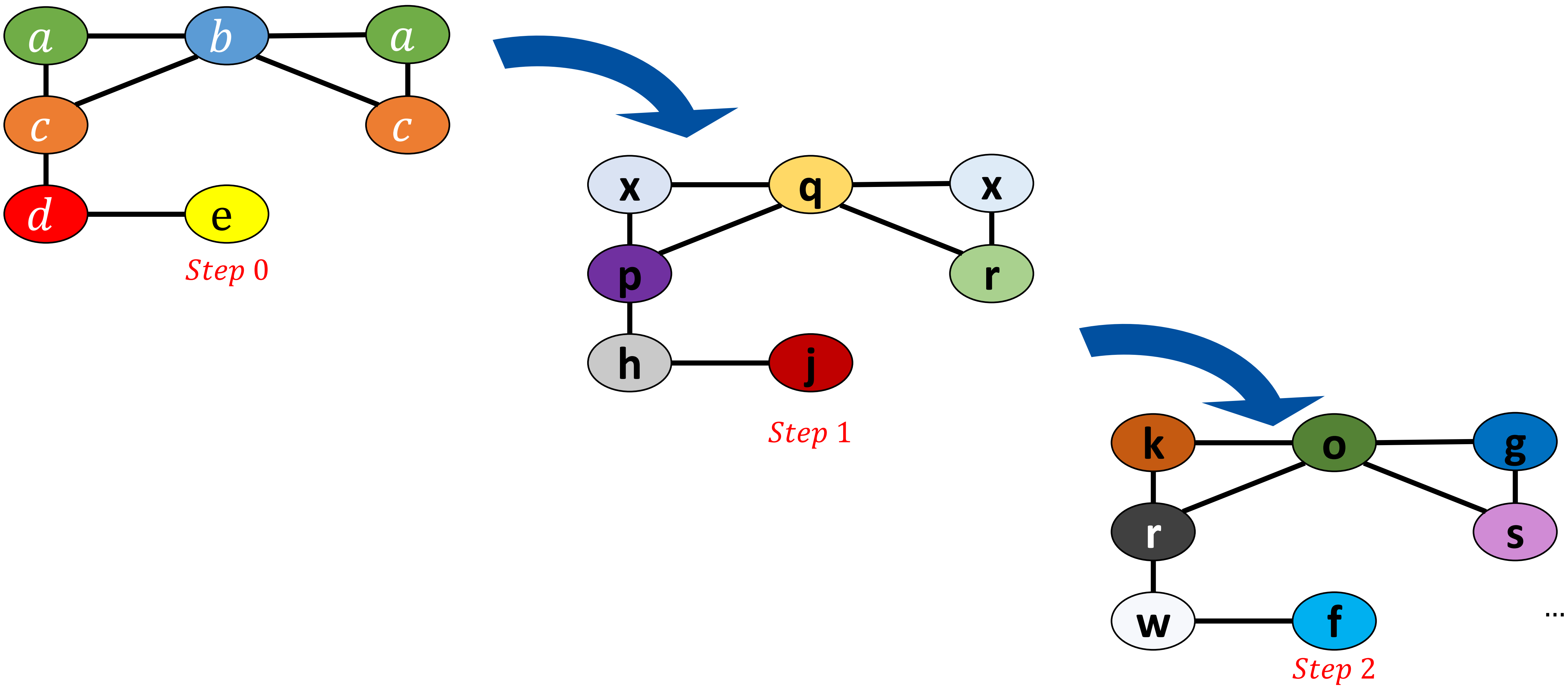
2.2 Graph Wasserstein distance
Motivation



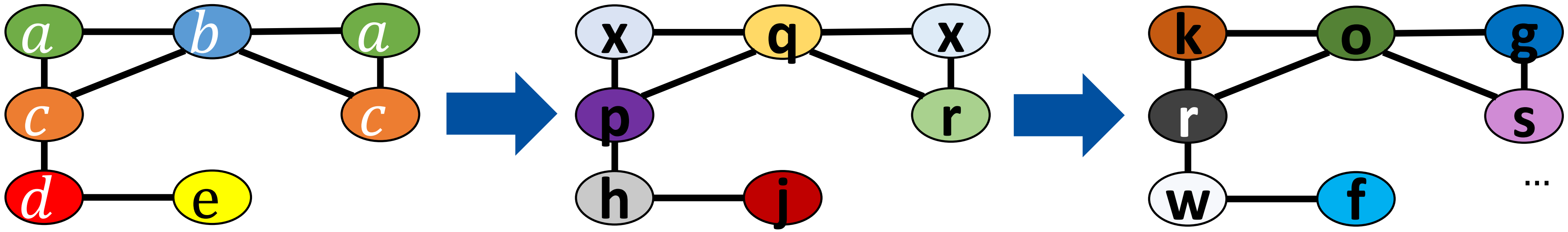
2.2 Graph Wasserstein distance
Motivation

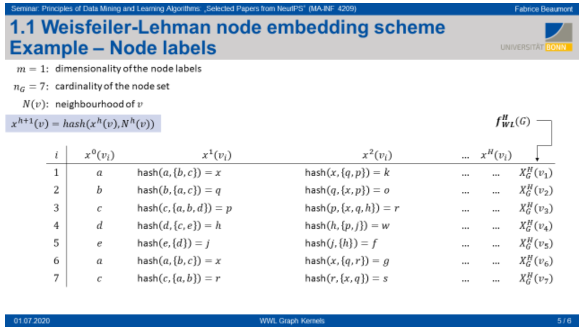


2.2 Graph Wasserstein distance
Idea



2.2 Graph Wasserstein distance
Idea




Seminar: Principles of Data Mining and Learning Algorithms: „Selected Papers from NeurIPS“ (MA-INF 4209) Fabrice Beaumont

1.1 Weisfeiler-Lehman node embedding scheme
Example – Node labels

$m = 1$: dimensionality of the node labels

$n_G = 7$: cardinality of the node set

$N(v)$: neighbourhood of v

$x^{h+1}(v) = hash(x^h(v), N^h(v))$

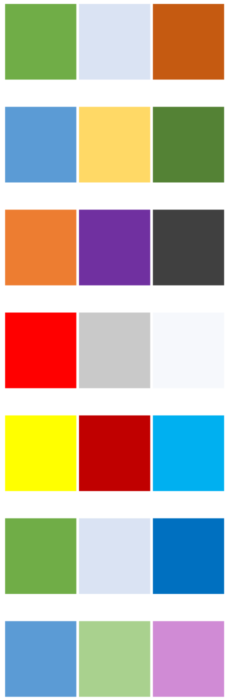
$f^H_{WL}(G)$

i	$x^0(v_i)$	$x^1(v_i)$	$x^2(v_i)$	...	$x^H(v_i)$	
1	a	$\text{hash}(a, \{b, c\}) = x$	$\text{hash}(x, \{q, p\}) = k$	...	...	$X_G^H(v_1)$
2	b	$\text{hash}(b, \{a, c\}) = q$	$\text{hash}(q, \{x, p\}) = o$	...	...	$X_G^H(v_2)$
3	c	$\text{hash}(c, \{a, b, d\}) = p$	$\text{hash}(p, \{x, q, h\}) = r$	...	...	$X_G^H(v_3)$
4	d	$\text{hash}(d, \{c, e\}) = h$	$\text{hash}(h, \{p, j\}) = w$	...	...	$X_G^H(v_4)$
5	e	$\text{hash}(e, \{d\}) = j$	$\text{hash}(j, \{h\}) = f$	...	...	$X_G^H(v_5)$
6	a	$\text{hash}(a, \{b, c\}) = x$	$\text{hash}(x, \{q, r\}) = g$	...	...	$X_G^H(v_6)$
7	c	$\text{hash}(c, \{a, b\}) = r$	$\text{hash}(r, \{x, q\}) = s$	...	...	$X_G^H(v_7)$

01.07.2020 WWL Graph Kernels 5 / 6

$(H = 2)$

$x^h(v_i)$



$v_1 \; = X_G^H\,(v_1)$

$v_2 \; = X_G^H\,(v_2)$

$v_3 \; = X_G^H\,(v_3)$

$v_4 \; = X_G^H\,(v_4)$

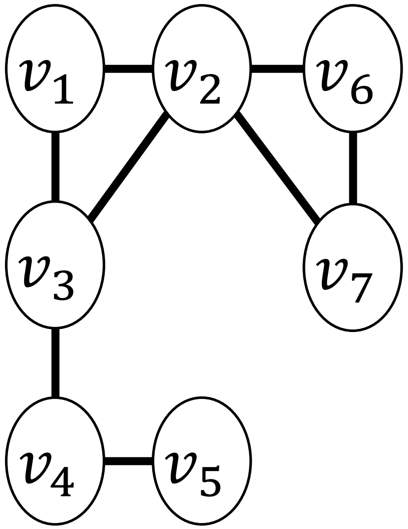
$v_5 \; = X_G^H\,(v_5)$

$v_6 \; = X_G^H\,(v_6)$

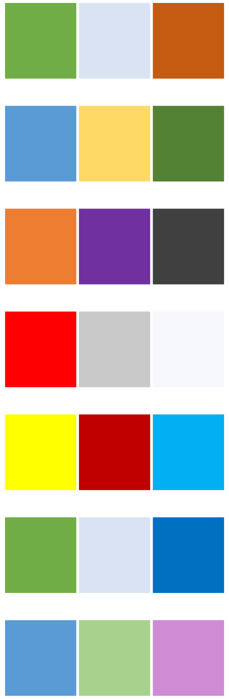
$v_7 \; = X_G^H\,(v_7)$

2.2 Graph Wasserstein distance
Idea

G



$x^h(v_i)$



$v_1 \quad = X_G^H\ (v_1)$

$v_2 \quad = X_G^H\ (v_2)$

$v_3 \quad = X_G^H\ (v_3)$

$v_4 \quad = X_G^H\ (v_4)$

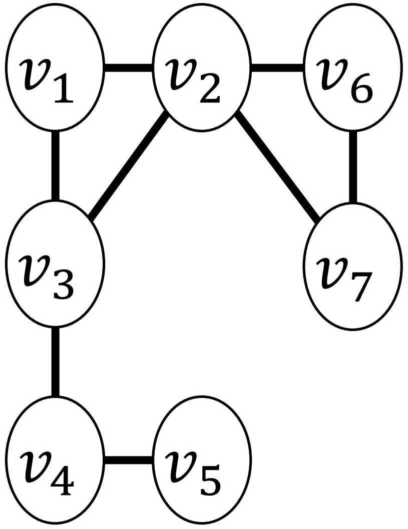
$v_5 \quad = X_G^H\ (v_5)$

$v_6 \quad = X_G^H\ (v_6)$

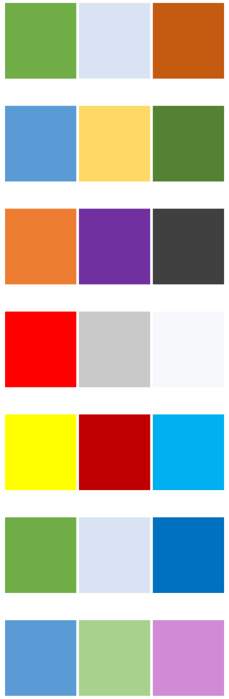
$v_7 \quad = X_G^H\ (v_7)$

2.2 Graph Wasserstein distance
Idea

G



$x^h(v_i)$



$v_1 \quad = X_G^H(v_1)$

$v_2 \quad = X_G^H(v_2)$

$v_3 \quad = X_G^H(v_3)$

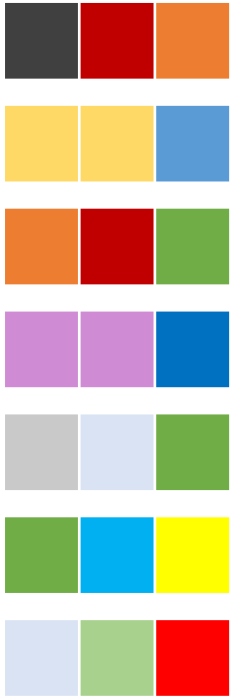
$v_4 \quad = X_G^H(v_4)$

$v_5 \quad = X_G^H(v_5)$

$v_6 \quad = X_G^H(v_6)$

$v_7 \quad = X_G^H(v_7)$

$x^h(w_i)$



$X_{G'}^H(w_1) = \quad w_1$

$X_{G'}^H(w_2) = \quad w_2$

$X_{G'}^H(w_3) = \quad w_3$

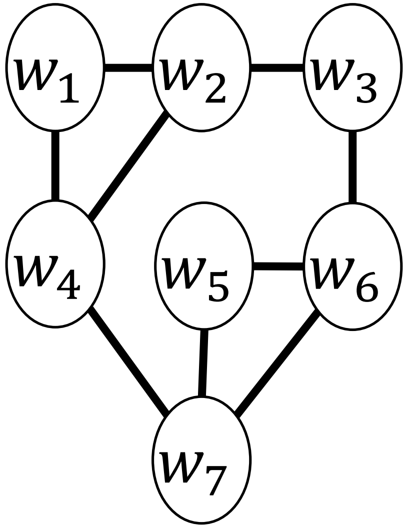
$X_{G'}^H(w_4) = \quad w_4$

$X_{G'}^H(w_5) = \quad w_5$

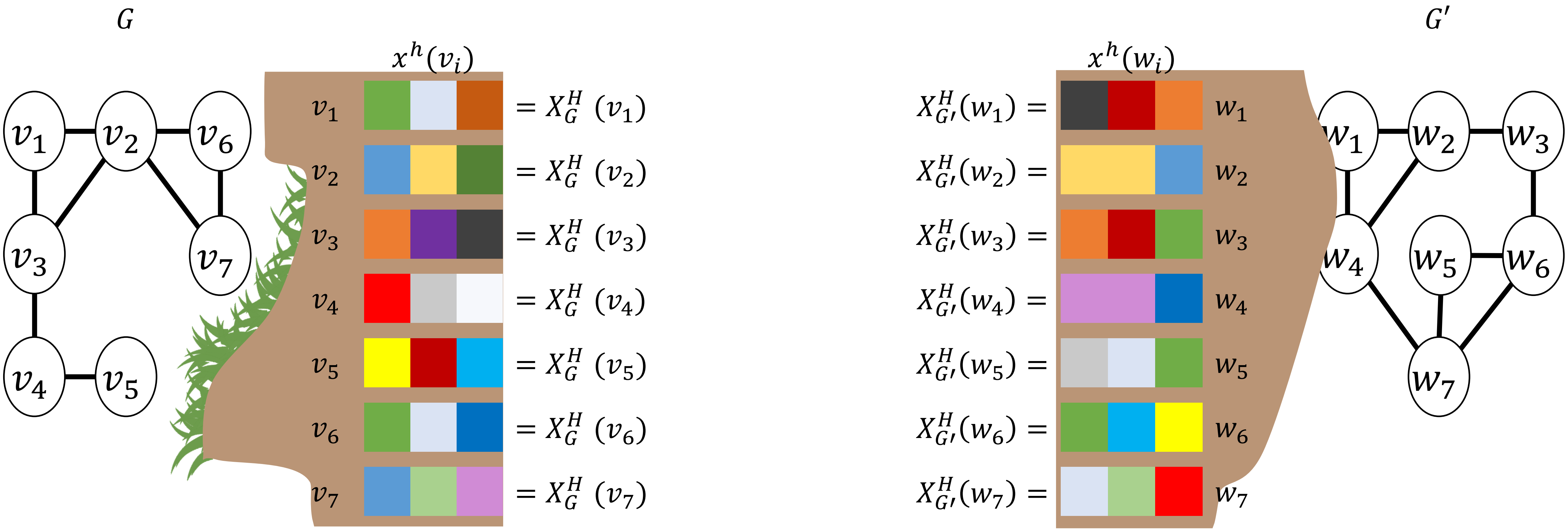
$X_{G'}^H(w_6) = \quad w_6$

$X_{G'}^H(w_7) = \quad w_7$

G'

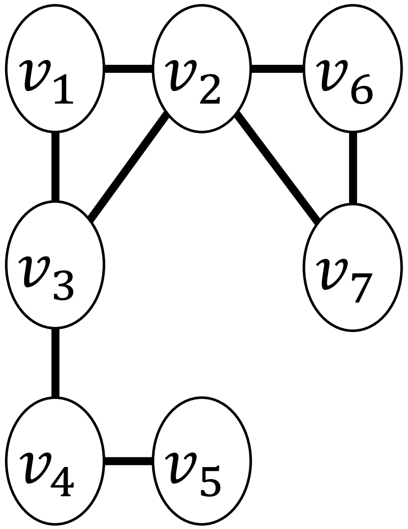


2.2 Graph Wasserstein distance
Idea

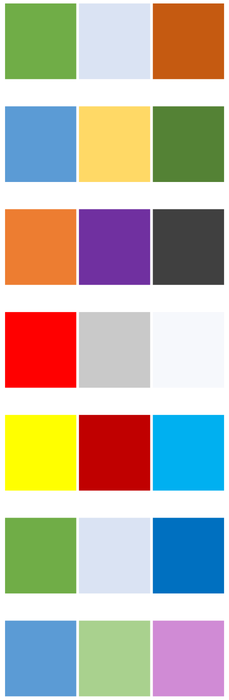


2.2 Graph Wasserstein distance
Transportation plan

G



$x^h(v_i)$



$v_1 = X_G^H(v_1)$

$v_2 = X_G^H(v_2)$

$v_3 = X_G^H(v_3)$

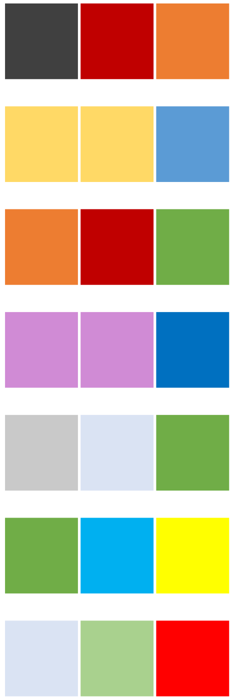
$v_4 = X_G^H(v_4)$

$v_5 = X_G^H(v_5)$

$v_6 = X_G^H(v_6)$

$v_7 = X_G^H(v_7)$

$x^h(w_i)$



$X_{G'}^H(w_1) = \; w_1$

$X_{G'}^H(w_2) = \; w_2$

$X_{G'}^H(w_3) = \; w_3$

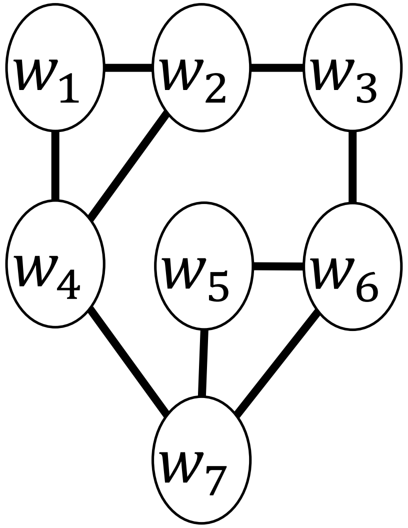
$X_{G'}^H(w_4) = \; w_4$

$X_{G'}^H(w_5) = \; w_5$

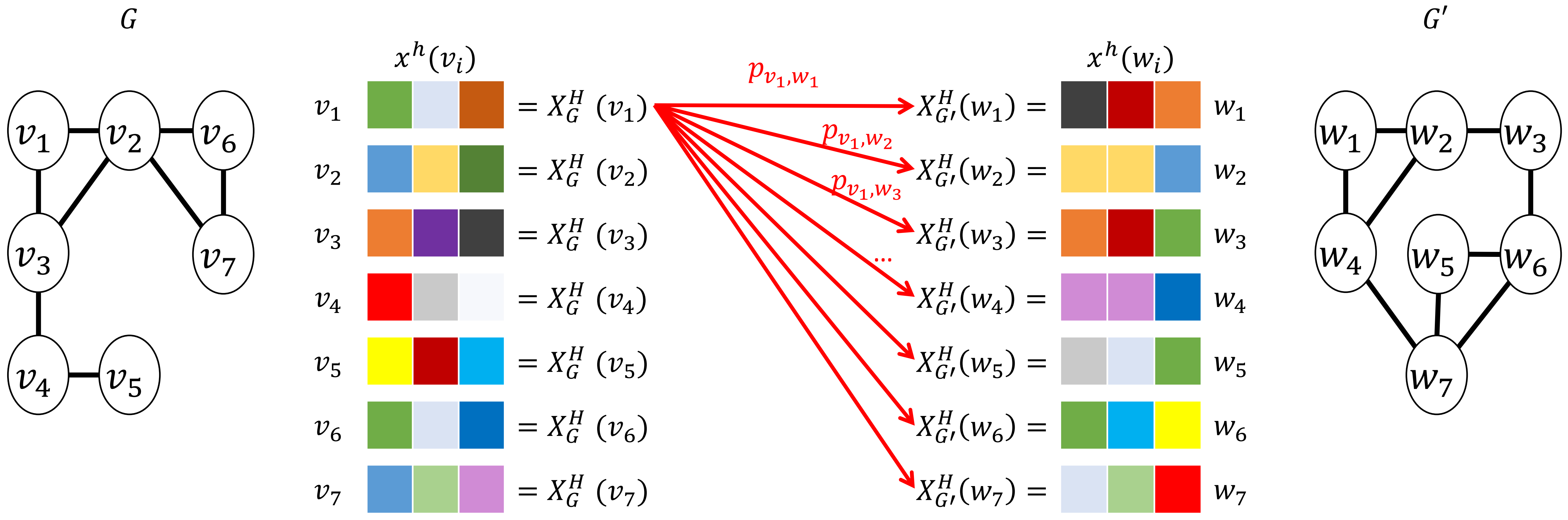
$X_{G'}^H(w_6) = \; w_6$

$X_{G'}^H(w_7) = \; w_7$

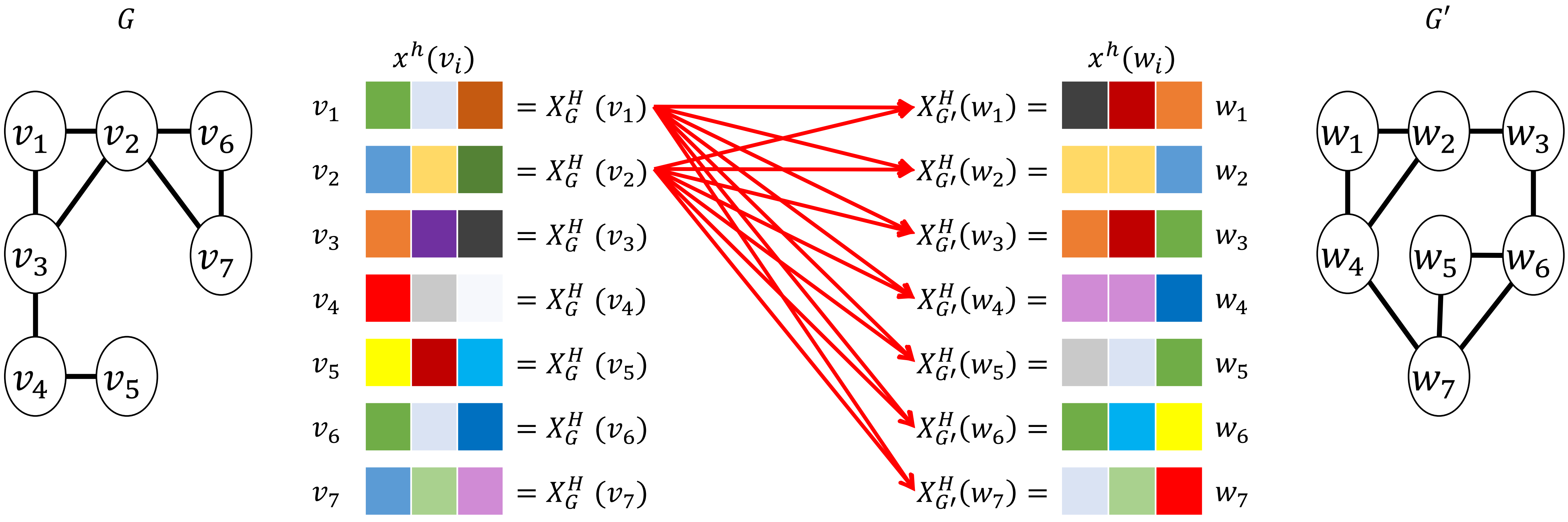
G'



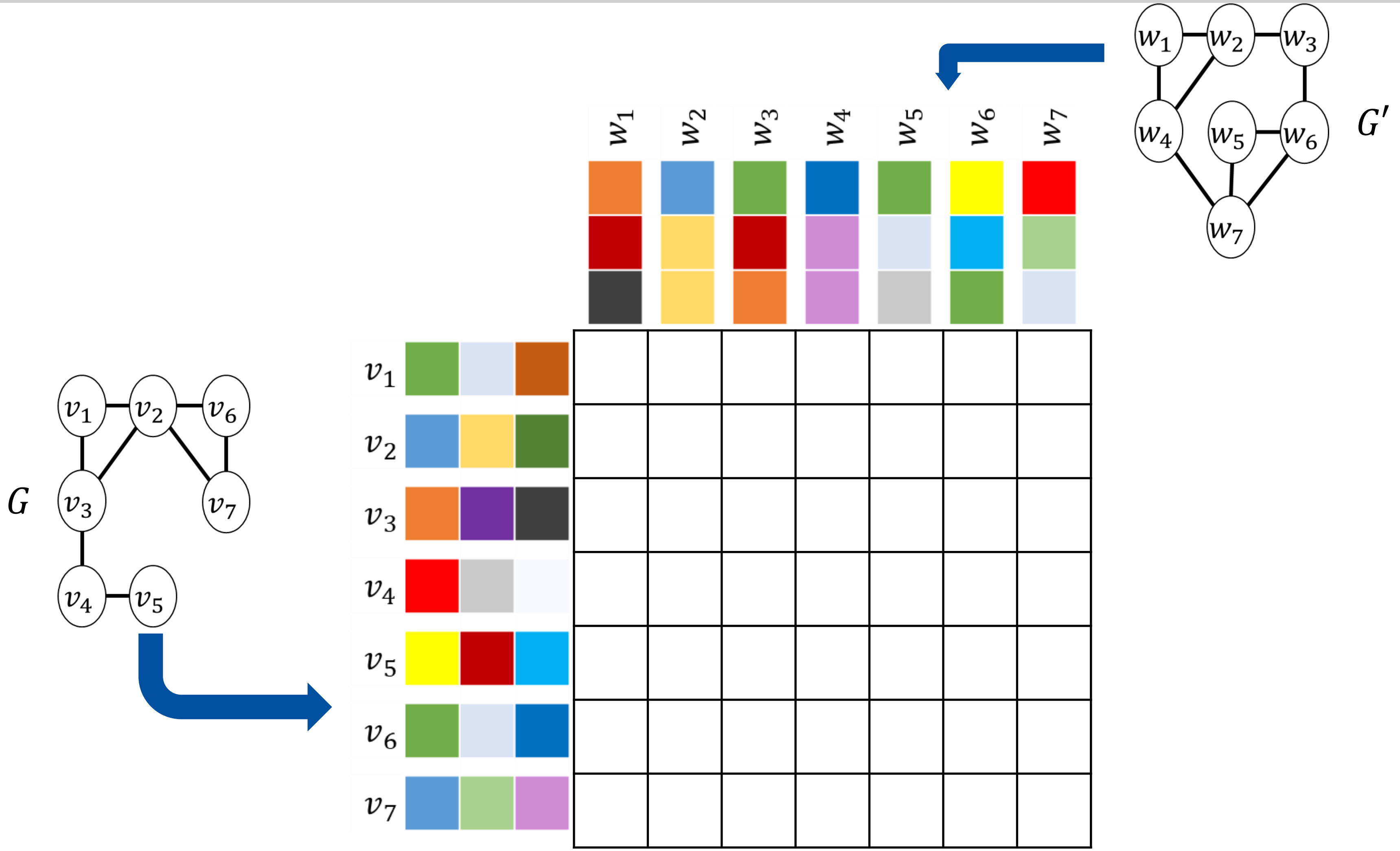
2.2 Graph Wasserstein distance
Transportation plan



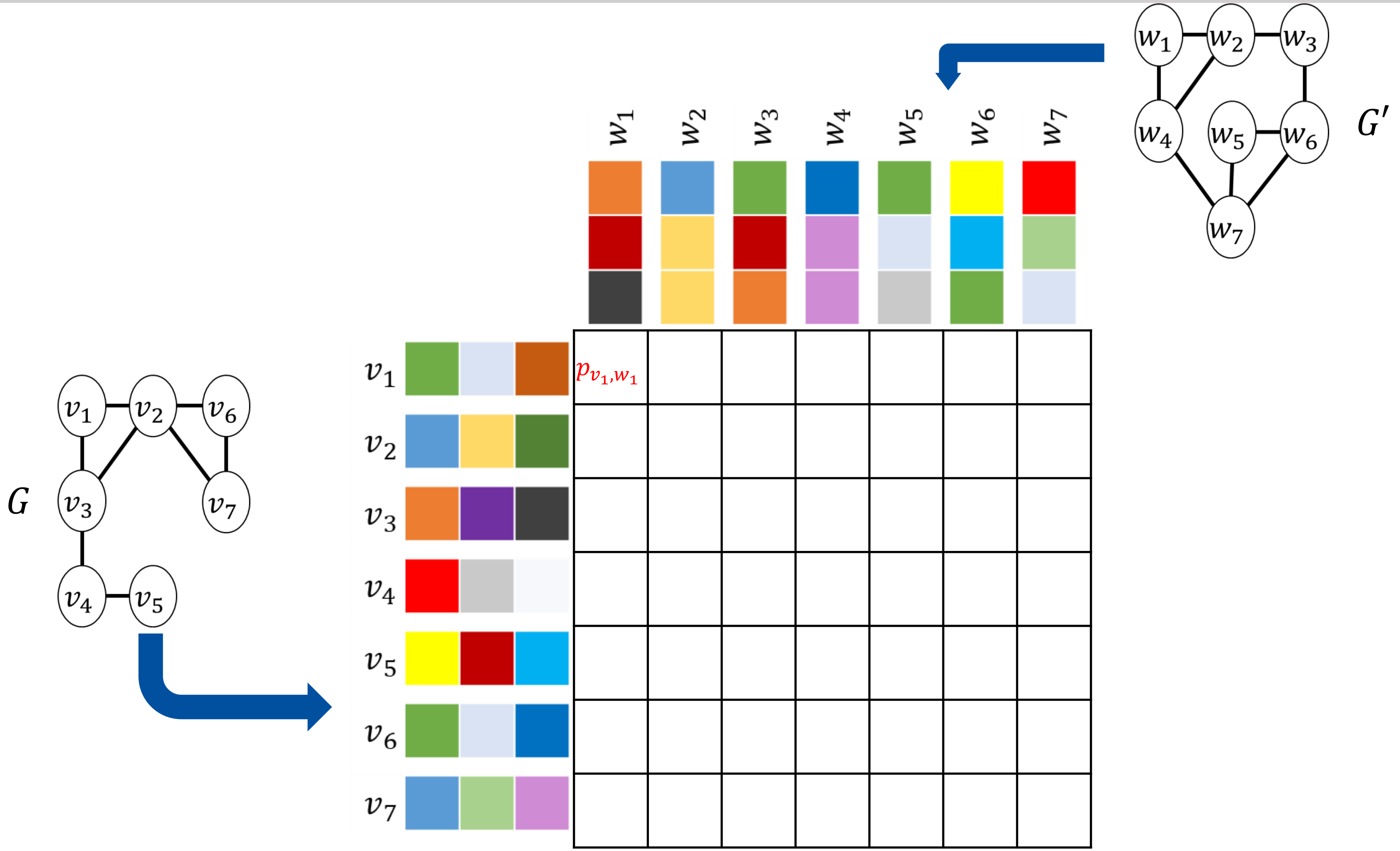
2.2 Graph Wasserstein distance
Transportation plan



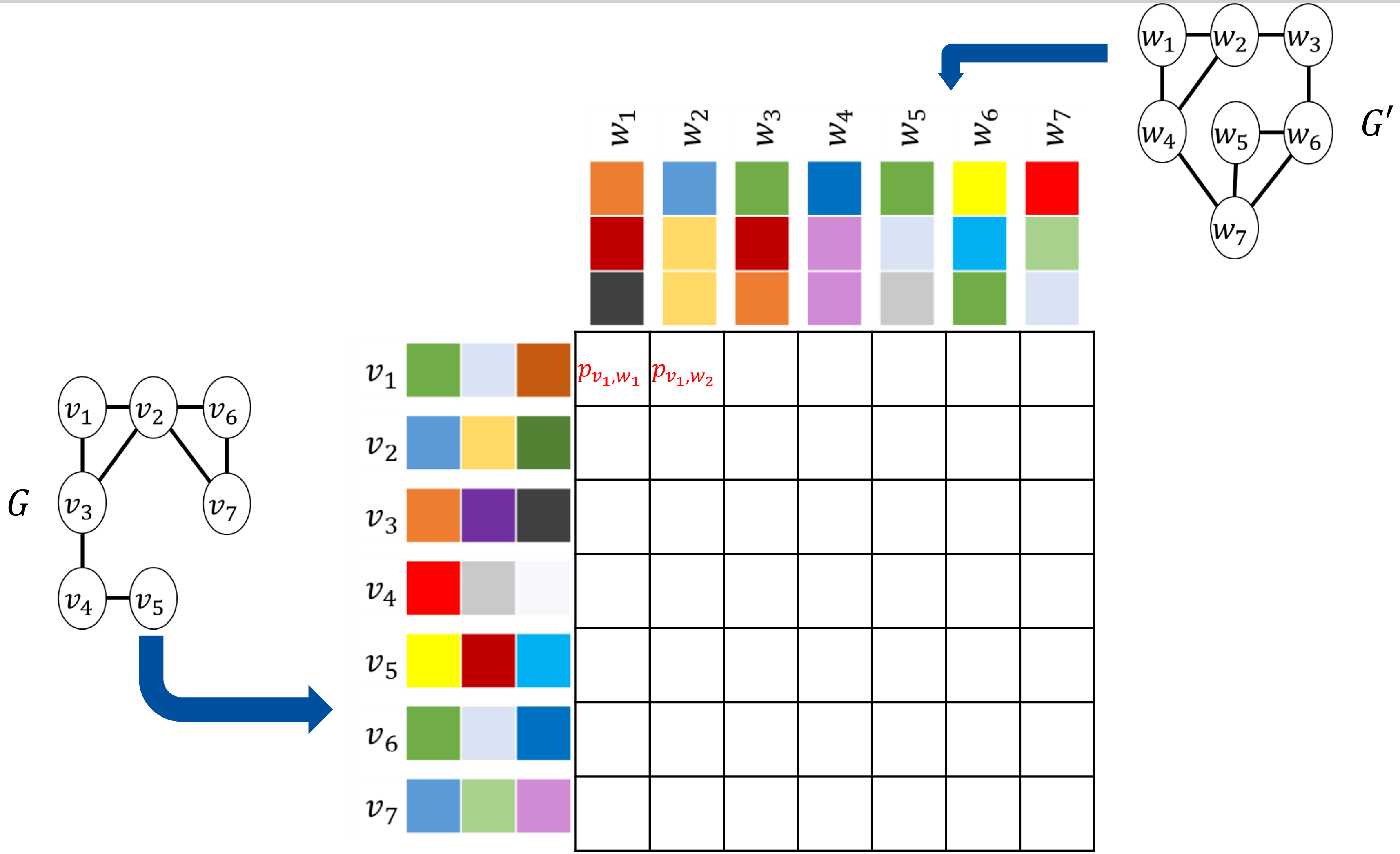
2.2 Graph Wasserstein distance
Transportation plan



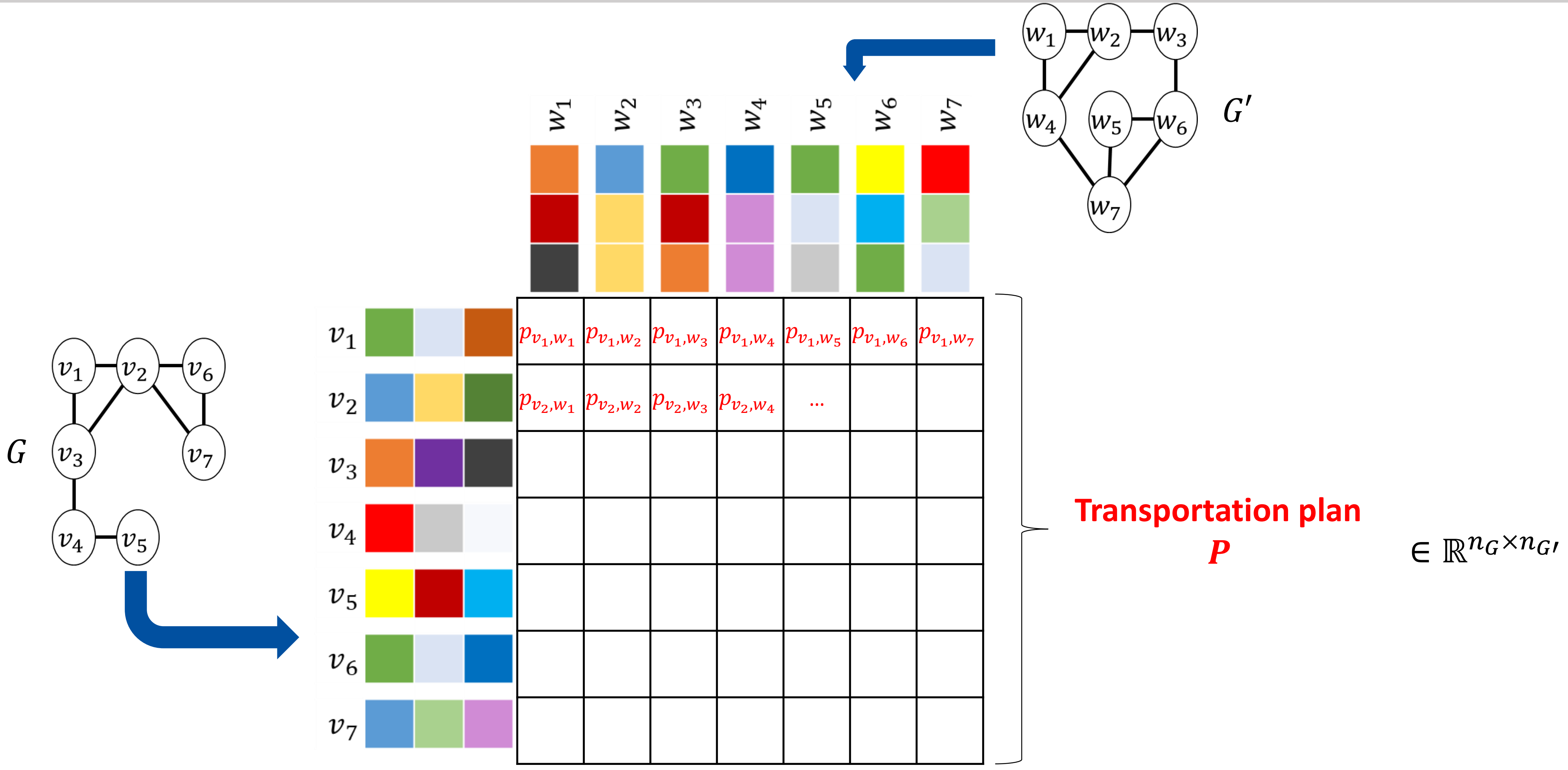
2.2 Graph Wasserstein distance
Transportation plan



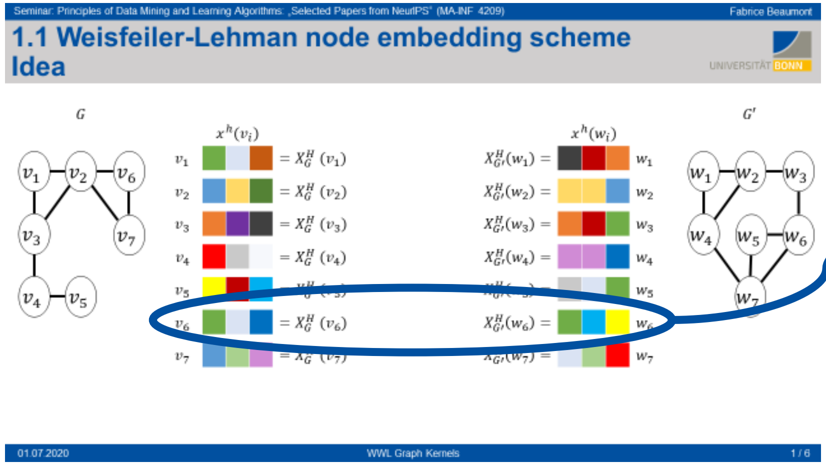
2.2 Graph Wasserstein distance
Transportation plan

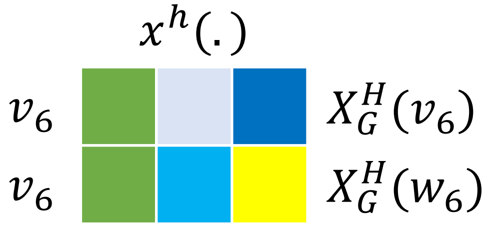


2.2 Graph Wasserstein distance
Transportation plan

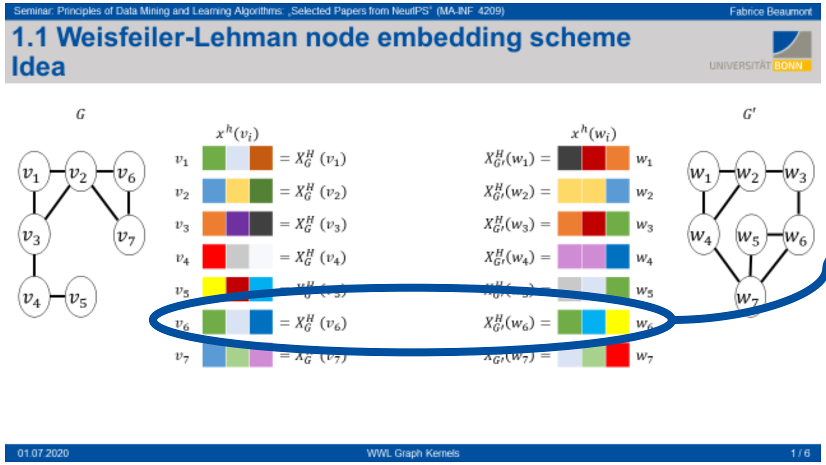


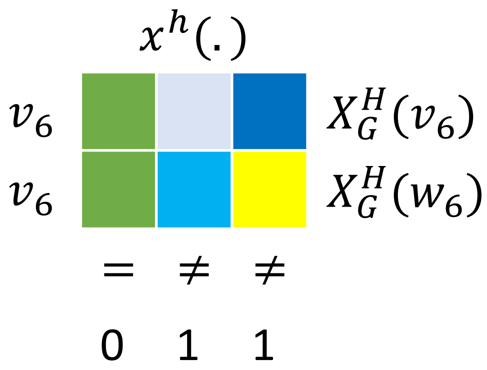
2.2 Graph Wasserstein distance
Distance measure



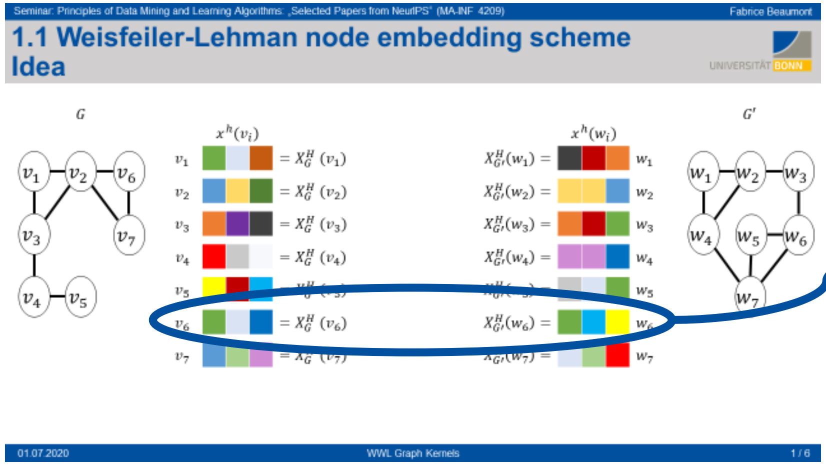


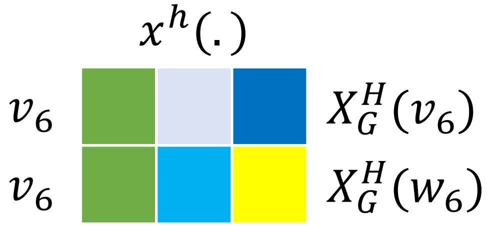
2.2 Graph Wasserstein distance
Distance measure





2.2 Graph Wasserstein distance
Distance measure



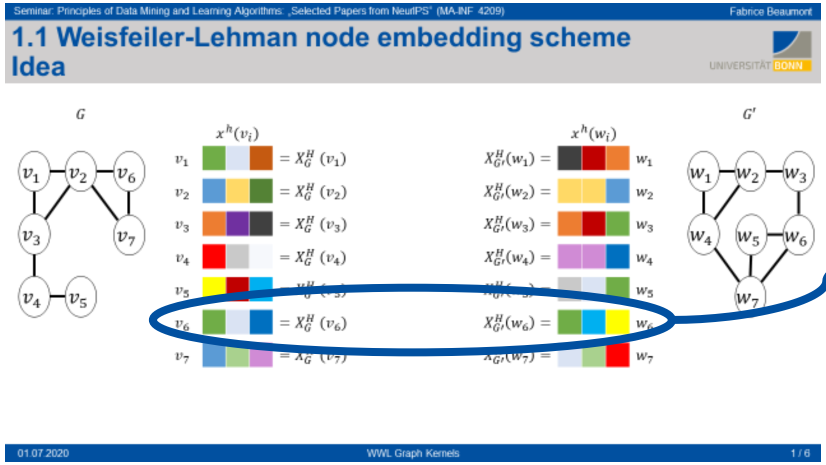


$=\quad \neq \quad \neq$

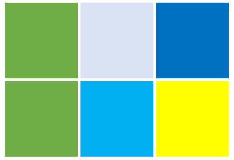
0 1 1

$$\Rightarrow d_{Ham} = \frac{1}{3}2$$

2.2 Graph Wasserstein distance
Distance measure



$x^h(.)$



$$= \quad \neq \quad \neq$$

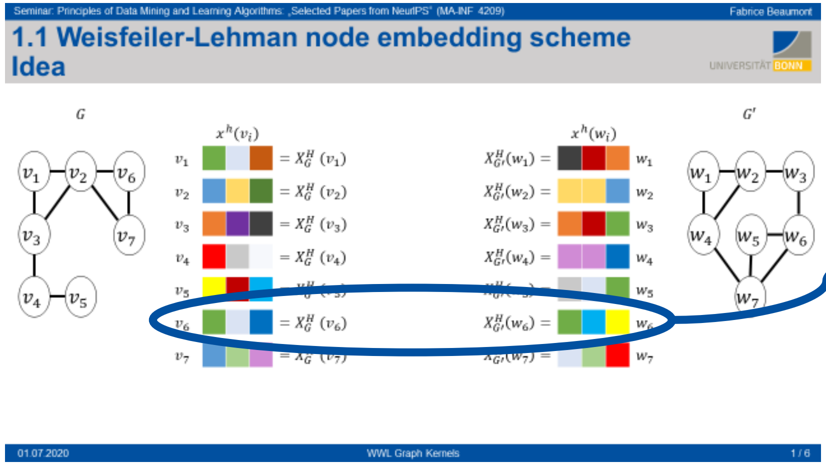
$$0 \quad 1 \quad 1$$

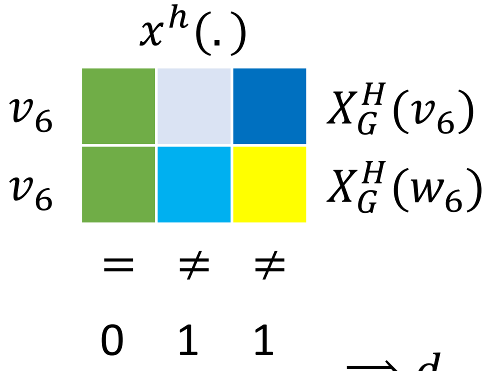
$$\Longrightarrow d_{Ham} = \frac{1}{3} 2$$

$$\rho(x,y) = \begin{cases} 1 \text{ if } x \neq y \\ 0 \text{ if } x = y \end{cases}$$

$$d_{Ham}\left(X_G^H(v_6), X_G^H(w_6)\right) = \frac{1}{H+1} \sum_{h=0}^{H} \rho(x^h(v_6), x^h(w_6))$$

2.2 Graph Wasserstein distance
Distance measure





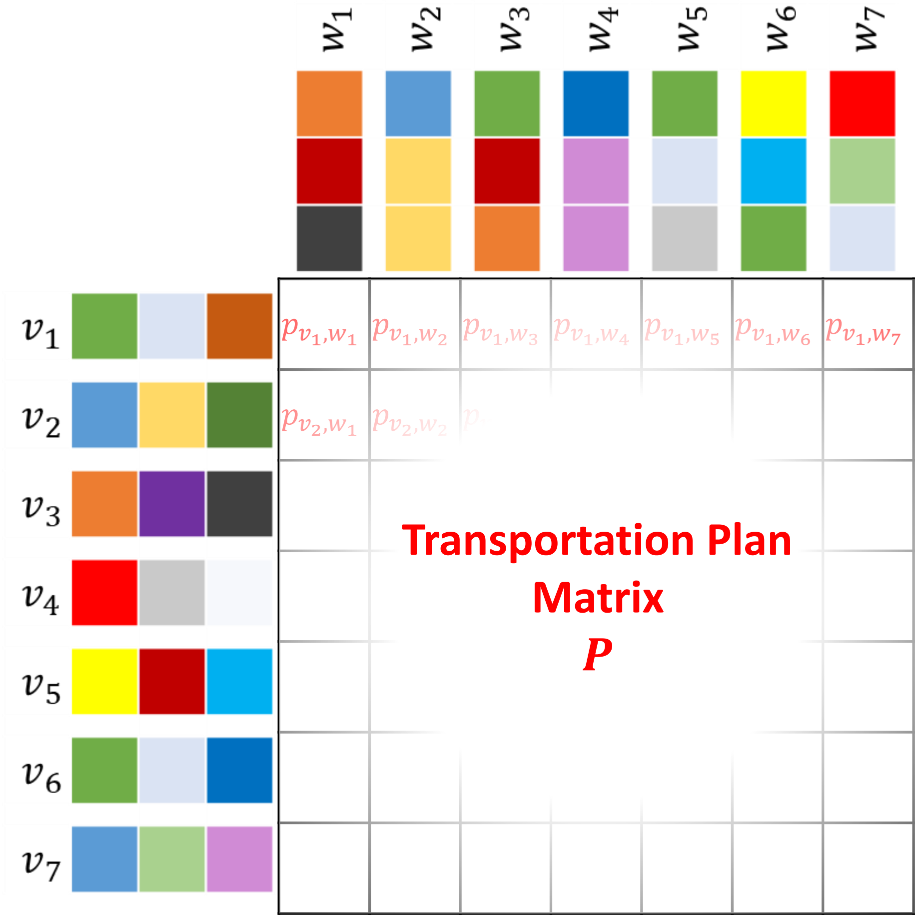
$$\Rightarrow d_{Ham} = \frac{1}{3}2$$

$$d_{Ham}\left(X_G^H(v_6), X_G^H(w_6)\right) = \frac{1}{H+1}\sum_{h=0}^{H}\rho(x^h(v_6), x^h(w_6))$$

$$\rho(x,y) = \begin{cases} 1 \text{ if } x \neq y \\ 0 \text{ if } x = y \end{cases}$$

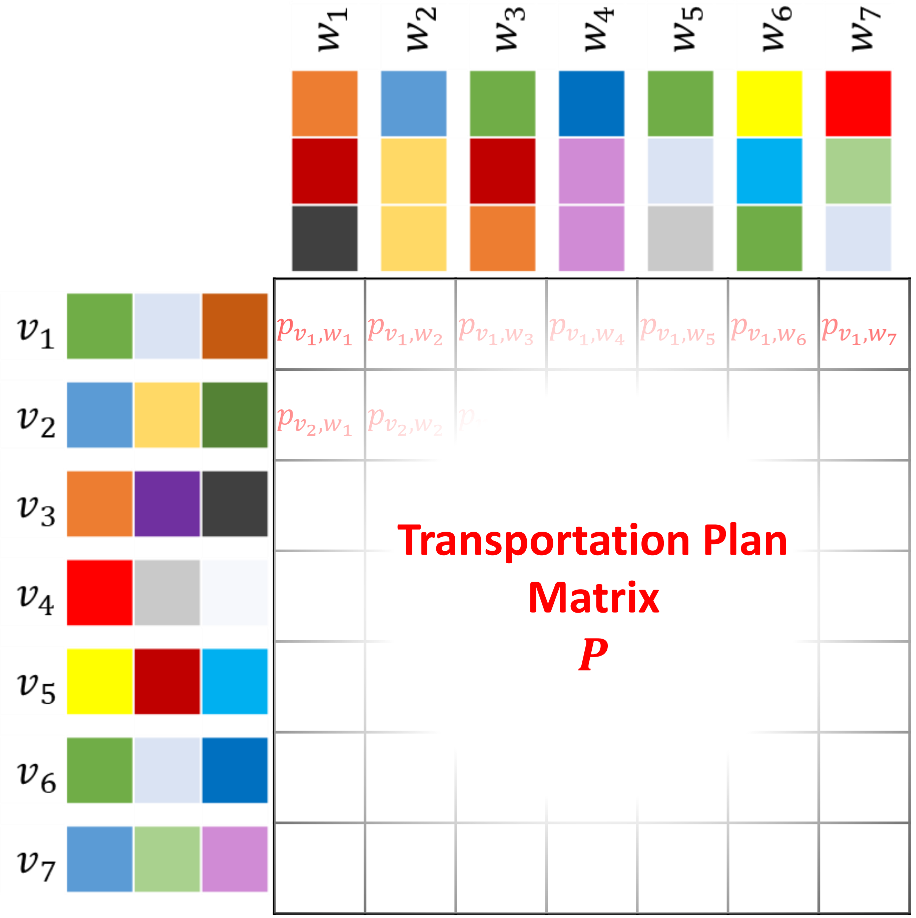
DISCRETE node labels	CONTINUOUS node labels
Normalised Hamming distance: $d_{Ham}(v,v') = \frac{1}{H+1}\sum_{i=0}^{H}\rho(v_i, v'_i)$	Euclidean distance: $d_E(v,v') = \lVert v - v' \rVert_2$

2. Graph Wasserstein distance

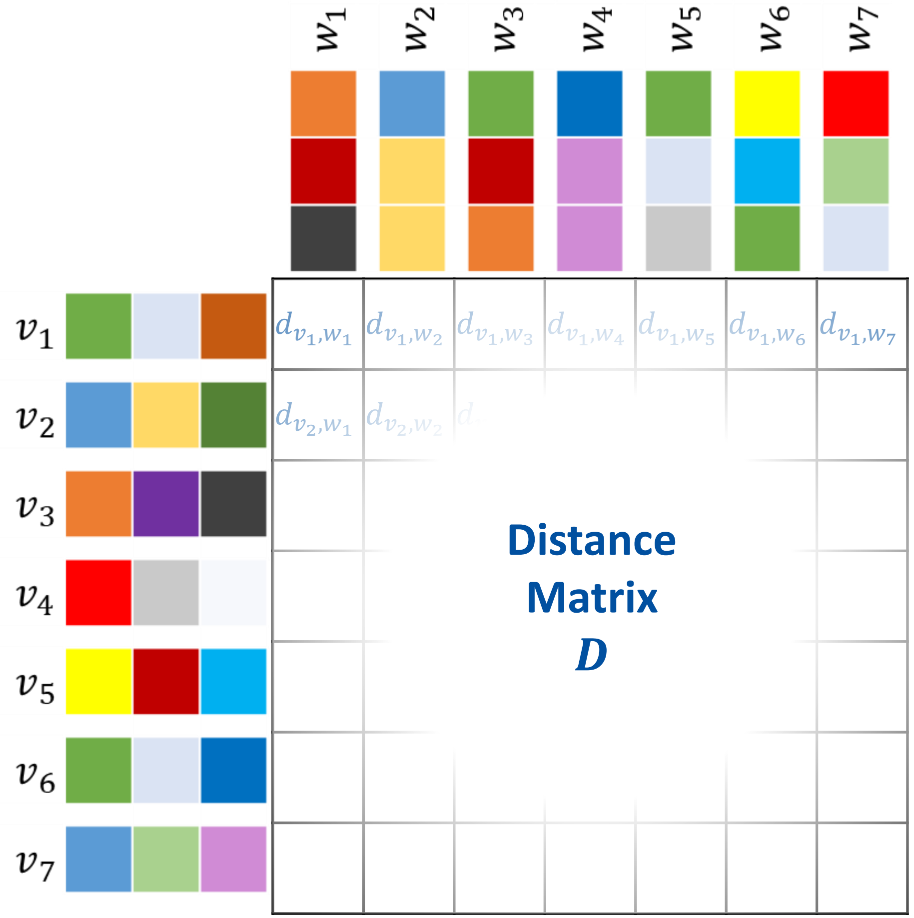


$\in \mathbb{R}^{n_G \times n_{G'}}$

2. Graph Wasserstein distance

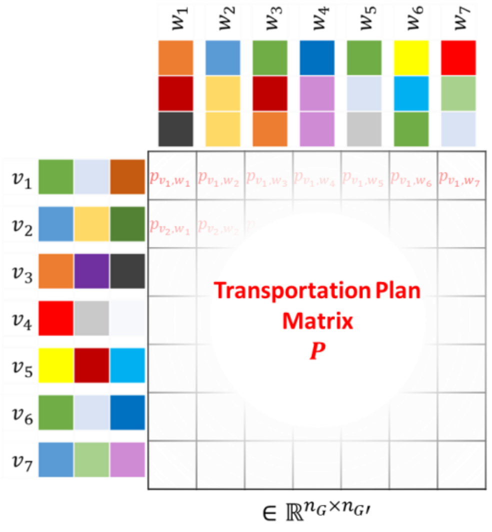


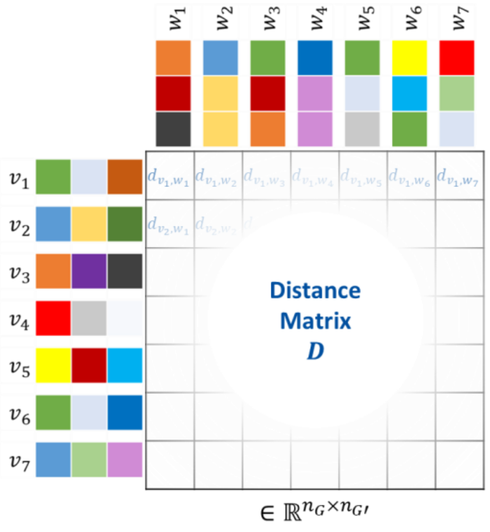
$\in \mathbb{R}^{n_G \times n_{G'}}$



$\in \mathbb{R}^{n_G \times n_{G'}}$

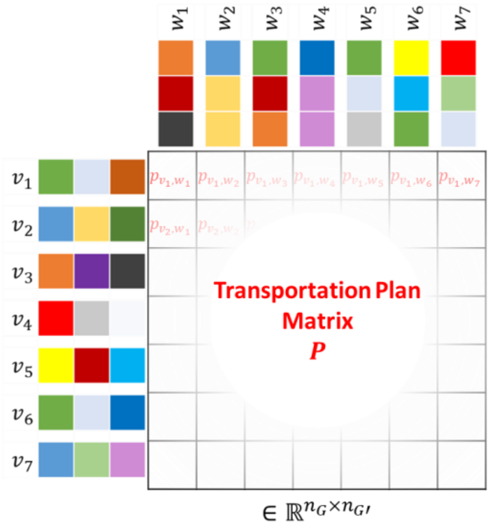
2. Graph Wasserstein distance

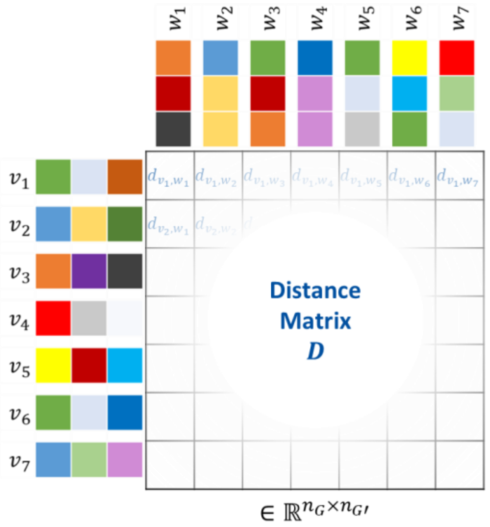




$$W_1(f(G), f(G')) := \min_{P} \sum_{\substack{v \in G, \\ w \in G'}} p_{v,w}\, d(v,w)$$

2. Graph Wasserstein distance





$$W_1(f(G), f(G')) := \min_P \sum_{\substack{v \in G, \\ w \in G'}} p_{v,w}\, d(v,w)$$

$$:= \min_{P \in \Gamma(f(G), f(G'))} \langle P, D \rangle$$

3 Kernel definition

Use a Laplacian kernel:

$$K_{WWL} = e^{-\lambda W_1(f(G), f(G'))}$$

(Leads to favourable conditions for the positive definiteness in case of non-Euclidean Distances)

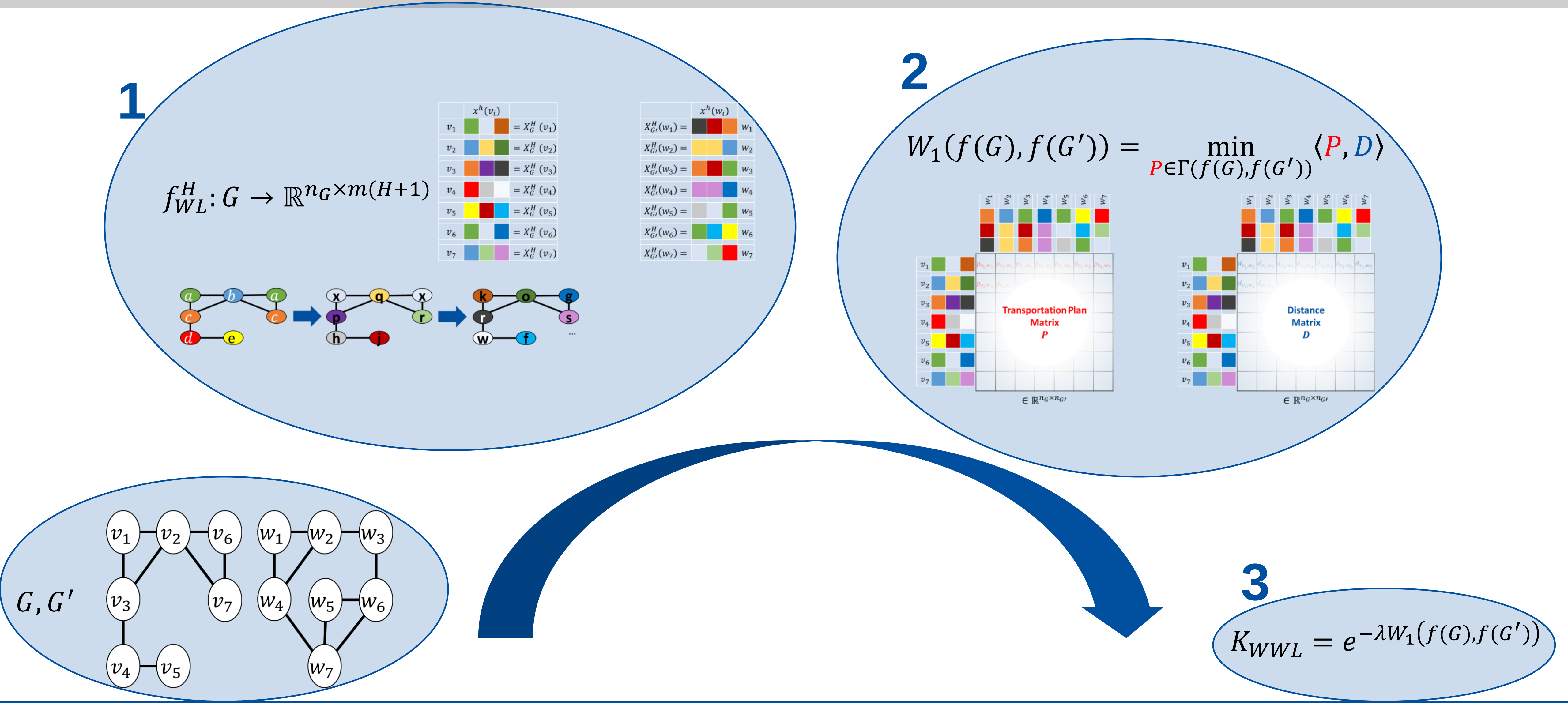
3 Kernel definition

DISCRETE node labels	CONTINUOUS node labels
• The kernel is positive definite for all $\lambda > 0$	? Open problem

3 Kernel definition

DISCRETE node labels	CONTINUOUS node labels
• Positive definite for all $\lambda > 0$ • SVM	• Positive definite? *Open problem* Workaround: • RKKS (reproducing kernel Krein spaces) • KSVM (Krein SVM)

4 OVERVIEW



1
$f_{WL}^H: G \to \mathbb{R}^{n_G \times m(H+1)}$
2
$W_1(f(G), f(G')) = \min_{P \in \Gamma(f(G), f(G'))} \langle P, D \rangle$
Transportation Plan Matrix P
Distance Matrix D
$\in \mathbb{R}^{n_G \times n_{G'}}$
G, G'
3
$K_{WWL} = e^{-\lambda W_1(f(G), f(G'))}$

5 Experimental evaluation

UNIVERSITÄT BONN

DISCRETE node labels

SVM accuracies [%], 10-fold cross-validation, trained parameters

	METHOD	MUTAG	PTC-MR	NCI1	PROTEINS	D&D	ENZYMES	
Vertex histogram	V	85.39±0.73	58.35±0.20	64.22±0.11	72.12±0.19	78.24±0.28	22.72±0.56	
Edge histogram	E	84.17±1.44	55.82±0.00	63.57±0.12	72.18±0.42	75.49±0.21	21.87±0.64	
Weisfeiler-Lehman kernel	WL	85.78±0.83	61.21±2.28	85.83±0.09	74.99±0.28	78.29±0.30	53.33±0.93	
Optimal assignment kernel	WL-OA	**87.15±1.82**	60.58±1.35	**86.08±0.27**	**76.37±0.30***	79.15±0.33	**58.97±0.82**	State of art
	WWL	**87.27±1.50**	**66.31±1.21***	85.75±0.25	74.28±0.56	**79.69±0.50**	**59.13±0.80**	

5 Experimental evaluation

DISCRETE node labels

SVM accuracies [%], 10-fold cross-validation, trained parameters

	METHOD	MUTAG	PTC-MR	NCI1	PROTEINS	D&D	ENZYMES
Vertex histogram	V	85.39±0.73	58.35±0.20	64.22±0.11	72.12±0.19	78.24±0.28	22.72±0.56
Edge histogram	E	84.17±1.44	55.82±0.00	63.57±0.12	72.18±0.42	75.49±0.21	21.87±0.64
Weisfeiler-Lehman kernel	WL	85.78±0.83	61.21±2.28	85.83±0.09	74.99±0.28	78.29±0.30	53.33±0.93
Optimal assignment kernel	WL-OA	**87.15±1.82**	60.58±1.35	**86.08±0.27**	**76.37±0.30***	79.15±0.33	**58.97±0.82**
	WWL	**87.27±1.50**	**66.31±1.21***	85.75±0.25	74.28±0.56	**79.69±0.50**	**59.13±0.80**

State of art

CONTINUOUS node labels

KSVM accuracies [%], 10-fold cross-validation, trained parameters

	METHOD	ENZYMES	PROTEINS	IMDB-B	BZR	COX2	BZR-MD	COX2-MD
Vertex histogram	VH-C	47.15±0.79	60.79±0.12	71.64±0.49	74.82±2.13	48.51±0.63	66.58±0.97	64.89±1.06
WWL with RBF-kernel	RBF-WL	68.43±1.47	75.43±0.28	72.06±0.34	80.96±1.67	75.45±1.53	**69.13±1.27**	71.83±1.61
Hash graph kernel	HGK-WL	63.04±0.65	75.93±0.17	73.12±0.40	78.59±0.63	**78.13±0.45**	**68.94±0.65**	74.61±1.74
	HGK-SP	66.36±0.37	75.78±0.17	73.06±0.27	76.42±0.72	72.57±1.18	66.17±1.05	68.52±1.00
GraphHopper kernel	GH	65.65±0.80	74.78±0.29	72.35±0.55	76.49±0.99	76.41±1.39	**69.14±2.08**	66.20±1.05
	WWL	**73.25±0.87***	**77.91±0.80***	**74.37±0.83***	**84.42±2.03***	**78.29±0.47**	**69.76±0.94**	**76.33±1.02**

New state of art

5 Experimental evaluation

DISCRETE node labels	CONTINUOUS node labels
• Competitive with best graph kernel	• Outperforms all state-of-the art graph kernels

Thank you

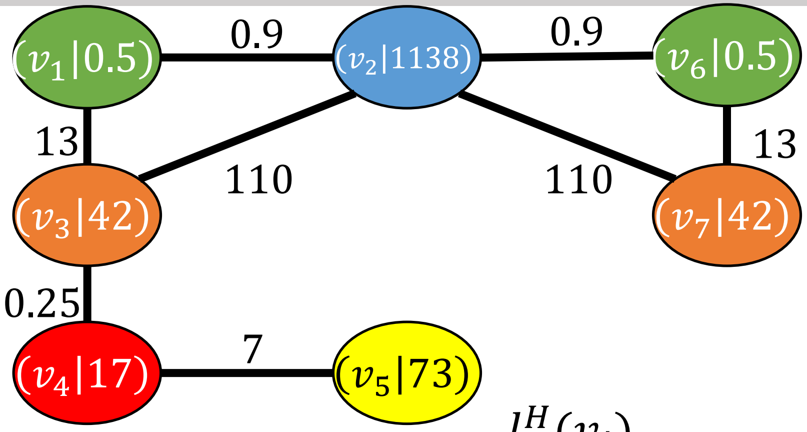
for your attention!

Appendix

Other slides

1.1 Weisfeiler-Lehman node embedding scheme Example – Attribute labels

$$a^{h+1}(v) = \frac{1}{2}\left(a^h(v) + \frac{1}{\deg(v)} \sum_{u \in N(v)} w((u,v))\, a^h(u) \right)$$

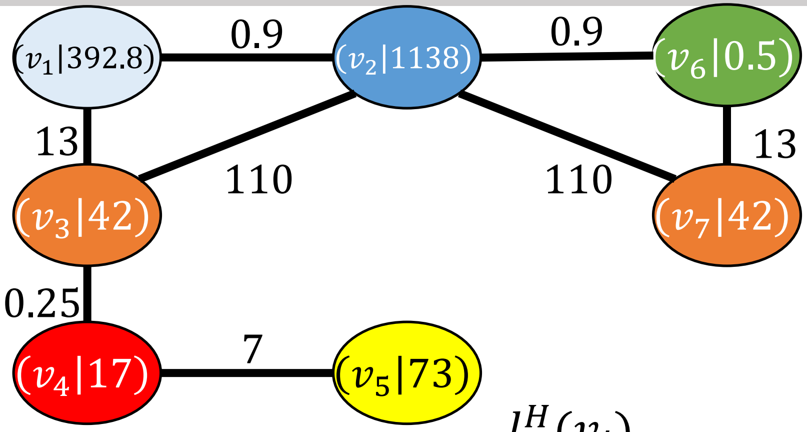


i	$l^0(v_i)$	$l^1(v_i)$	...	$l^H(v_i)$
1	0.5		...	...
2	42		...	...
3	1138		...	...
4	17		...	...
5	73		...	...
6	0.5		...	...
7	42		...	...

1.1 Weisfeiler-Lehman node embedding scheme
Example – Attribute labels

UNIVERSITÄT BONN

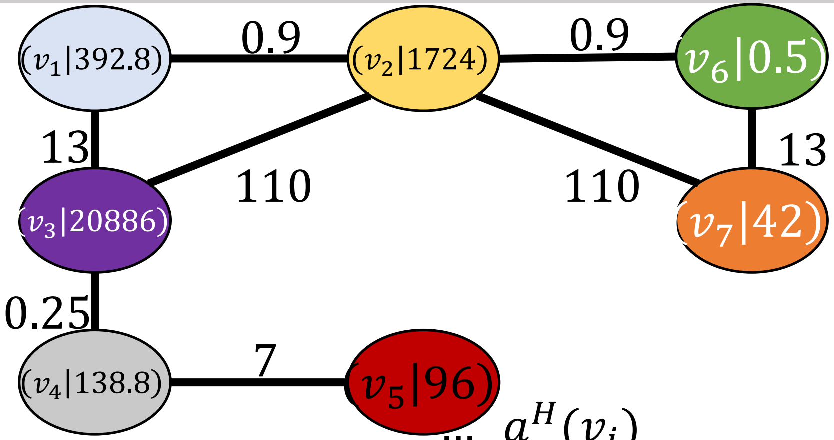
$$a^{h+1}(v) = \frac{1}{2}\left(a^h(v) + \frac{1}{\deg(v)} \sum_{u \in N(v)} w((u,v))\, a^h(u)\right)$$



i	$l^0(v_i)$	$l^1(v_i)$	…	$l^H(v_i)$
1	0.5	$^1/_2\,(0.5 + {}^1/_2\,[0.9 * 1138 + 13 * 42]) = 392.8$	…	…
2	42		…	…
3	1138		…	…
4	17		…	…
5	73		…	…
6	0.5		…	…
7	42		…	…

1.1 Weisfeiler-Lehman node embedding scheme
Example – Attribute labels

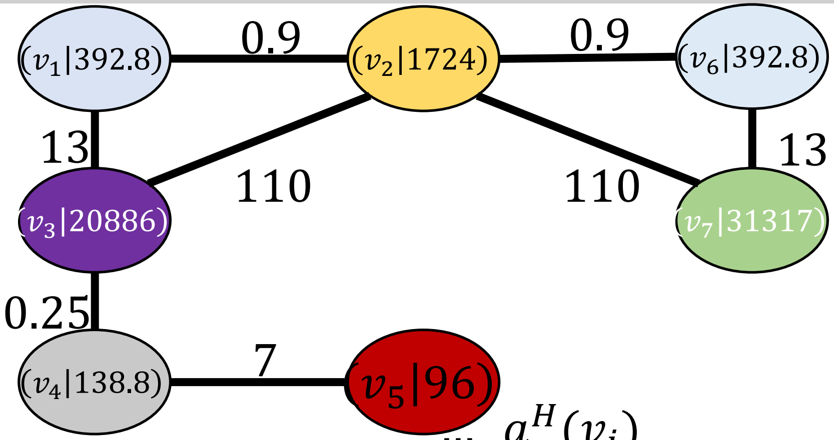
$$a^{h+1}(v) = \frac{1}{2}\left(a^h(v) + \frac{1}{\deg(v)} \sum_{u \in N(v)} w((u,v))\, a^h(u) \right)$$



i	$a^0(v_i)$	$a^1(v_i)$	...	$a^H(v_i)$
1	0.5	$\frac{1}{2}\left(0.5 + \frac{1}{2}[0.9 * 1138 + 13 * 42]\right) = 392.8$	...	...
2	42	$\frac{1}{2}\left(42 + \frac{1}{3}[13 * 0.5 + 110 * 1138 + 0.25 * 17]\right) = 20886.125$	...	...
3	1138	$\frac{1}{2}\left(1138 + \frac{1}{4}[0.9 * 0.5 + 110 * 42 + 0.9 * 0.5 + 110 * 42]\right) = 1724.1125$	...	...
4	17	$\frac{1}{2}\left(17 + \frac{1}{2}[0.25 * 42 + 7 * 73]\right) = 138.875$	...	...
5	73	$\frac{1}{2}\left(73 + \frac{1}{1}[7 * 17]\right) = 96$	...	...
6	0.5		...	...
7	42		...	...

1.1 Weisfeiler-Lehman node embedding scheme Example – Attribute labels

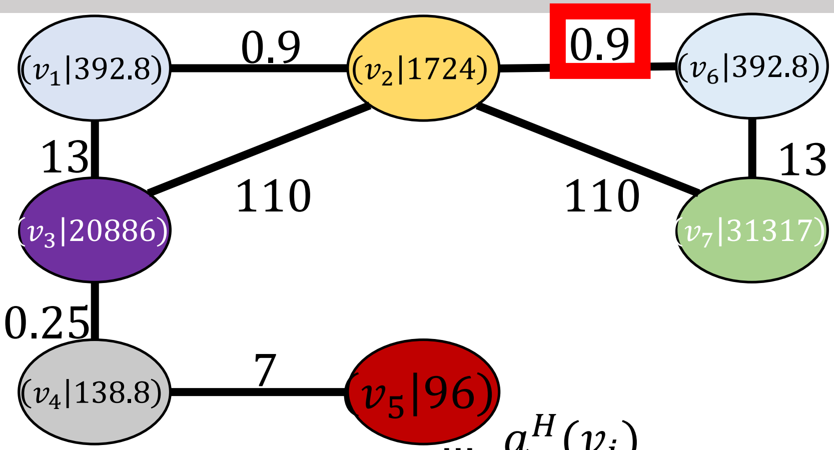
$$a^{h+1}(v) = \frac{1}{2}\left(a^h(v) + \frac{1}{\deg(v)} \sum_{u \in N(v)} w((u,v))\, a^h(u)\right)$$



i	$a^0(v_i)$	$a^1(v_i)$	…	$a^H(v_i)$
1	0.5	$^1/_2\left(0.5 + {^1/_2}\,[0.9 * 1138 + 13 * 42]\right) = 392.8$	…	…
2	42	$^1/_2\left(42 + {^1/_3}\,[13 * 0.5 + 110 * 1138 + 0.25 * 17]\right) = 20886.125$	…	…
3	1138	$^1/_2\left(1138 + {^1/_4}\,[0.9 * 0.5 + 110 * 42 + 0.9 * 0.5 + 110 * 42]\right) = 1724.1125$	…	…
4	17	$^1/_2\left(17 + {^1/_2}\,[0.25 * 42 + 7 * 73]\right) = 138.875$	…	…
5	73	$^1/_2\left(73 + {^1/_1}\,[7 * 17]\right) = 96$	…	…
6	0.5	$^1/_2\left(0.5 + {^1/_2}\,[0.9 * 1138 + 13 * 42]\right) = 392.8$	…	…
7	42	$^1/_2\left(42 + {^1/_2}\,[13 * 0.5 + 110 * 1138]\right) = 31317.625$	…	…

1.1 Weisfeiler-Lehman node embedding scheme Example – Attribute labels

$$a^{h+1}(v) = \frac{1}{2}\left(a^h(v) + \frac{1}{\deg(v)} \sum_{u \in N(v)} w((u,v))\, a^h(u)\right)$$



i	$a^0(v_i)$	$a^1(v_i)$	…	$a^H(v_i)$
1	0.5	$^1/_2\left(0.5 + {}^1/_2\,[0.9 * 1138 + 13 * 42]\right) = 392.8$	…	…
2	42	$^1/_2\left(42 + {}^1/_3\,[13 * 0.5 + 110 * 1138 + 0.25 * 17]\right) = 20886.125$	…	…
3	1138	$^1/_2\left(1138 + {}^1/_4\,[0.9 * 0.5 + 110 * 42 + \boxed{0.9} * 0.5 + 110 * 42]\right) = 1724.1125$	…	…
4	17	$^1/_2\left(17 + {}^1/_2\,[0.25 * 42 + 7 * 73]\right) = 138.875$	…	…
5	73	$^1/_2\left(73 + {}^1/_1\,[7 * 17]\right) = 96$	…	…
6	0.5	$^1/_2\left(0.5 + {}^1/_2\,[\boxed{0.9} * 1138 + 13 * 42]\right) = 392.8$	…	…
7	42	$^1/_2\left(42 + {}^1/_2\,[13 * 0.5 + 110 * 1138]\right) = 31317.625$	…	…

2.2 Wasserstein distance - L^p

$p \in [1, \infty)$: parameter for the $\boldsymbol{L^p}$ Wasserstein distance

σ, μ: probability distributions on a metric space $\boldsymbol{M}$

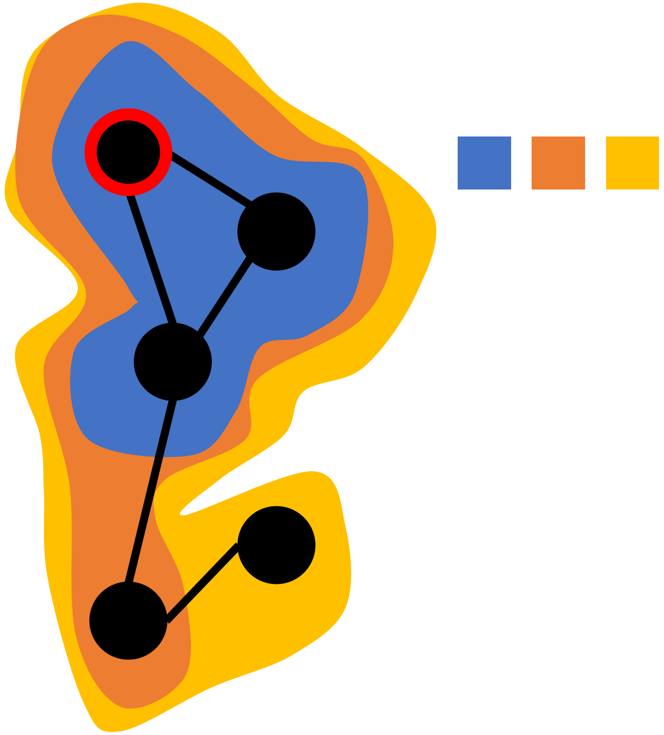
$d: \mathbb{R}^m x \mathbb{R}^m \rightarrow \mathbb{R}^m$: ground distance (on M)

$\Gamma(\sigma, \mu)$: set of all transportaion plans over MxM (with marginals σ and μ on the first and second factor)

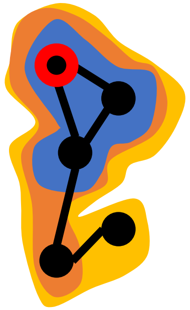
$$W_p(\sigma, \mu) := \left(\lim_{\gamma \in \Gamma(\sigma, \mu)} \int_{MxM} d(x, y)^p d\gamma(x, y) \right)^{\frac{1}{p}}$$

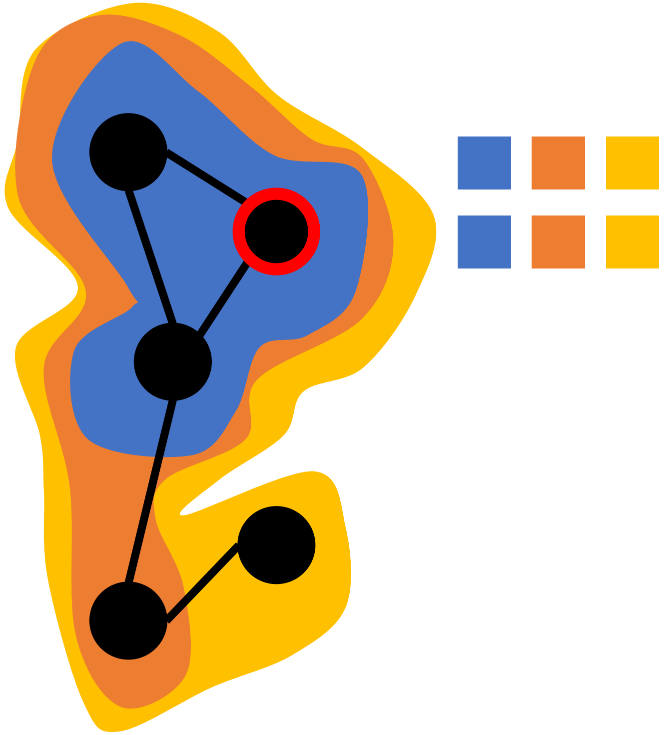
1.1 Weisfeiler-Lehman node embedding scheme
Idea

TODO: Fix this example: use different labels for the Node.
Mention formula next to example

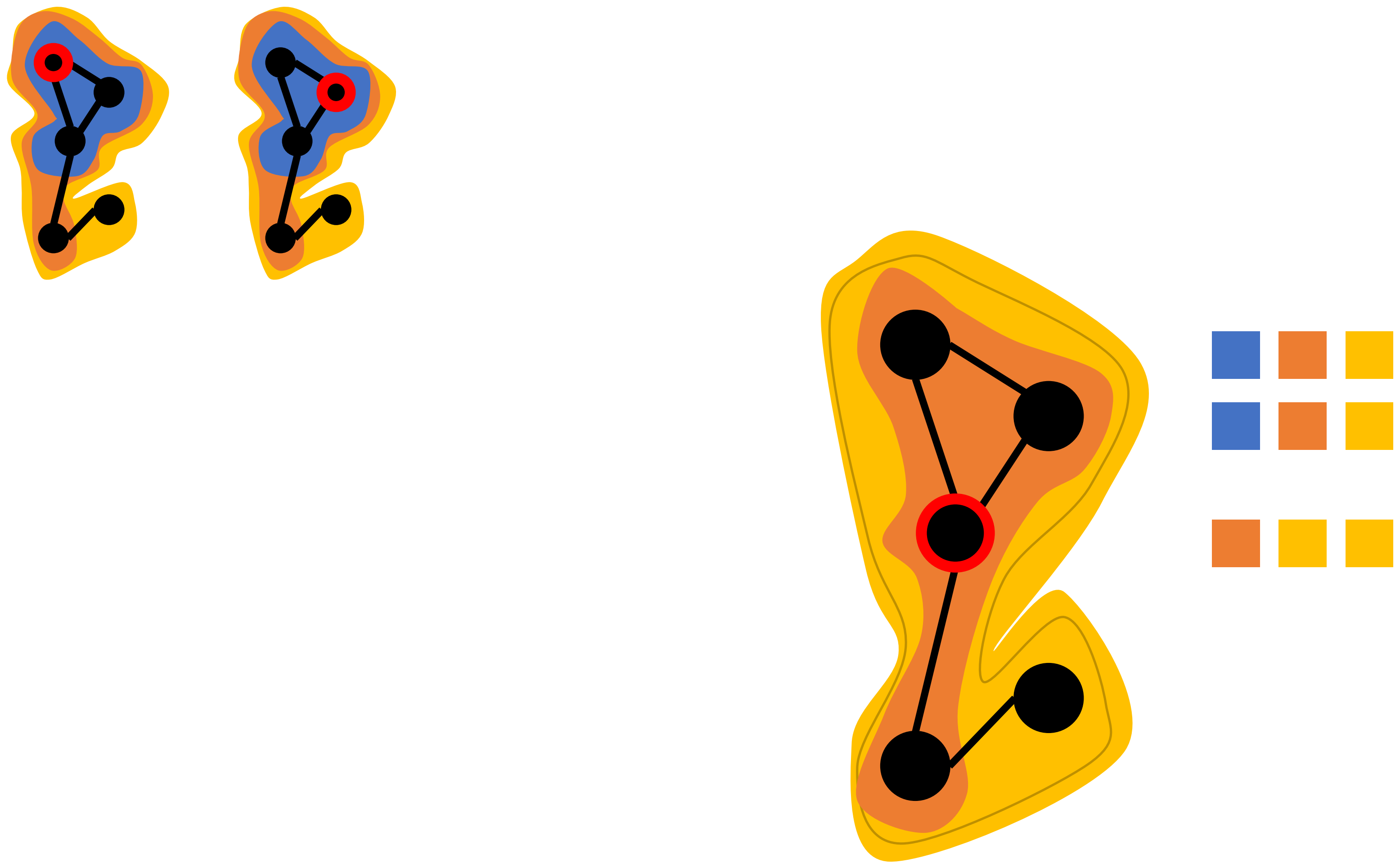


1.1 Weisfeiler-Lehman node embedding scheme
Idea

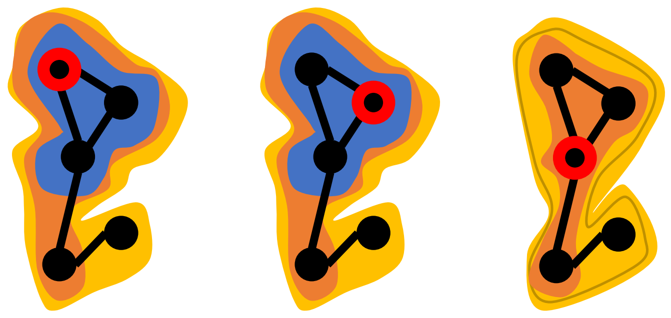


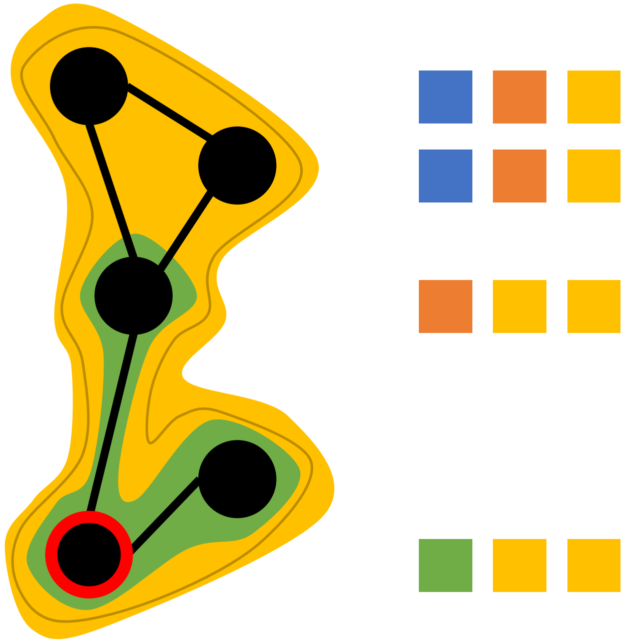


1.1 Weisfeiler-Lehman node embedding scheme
Idea

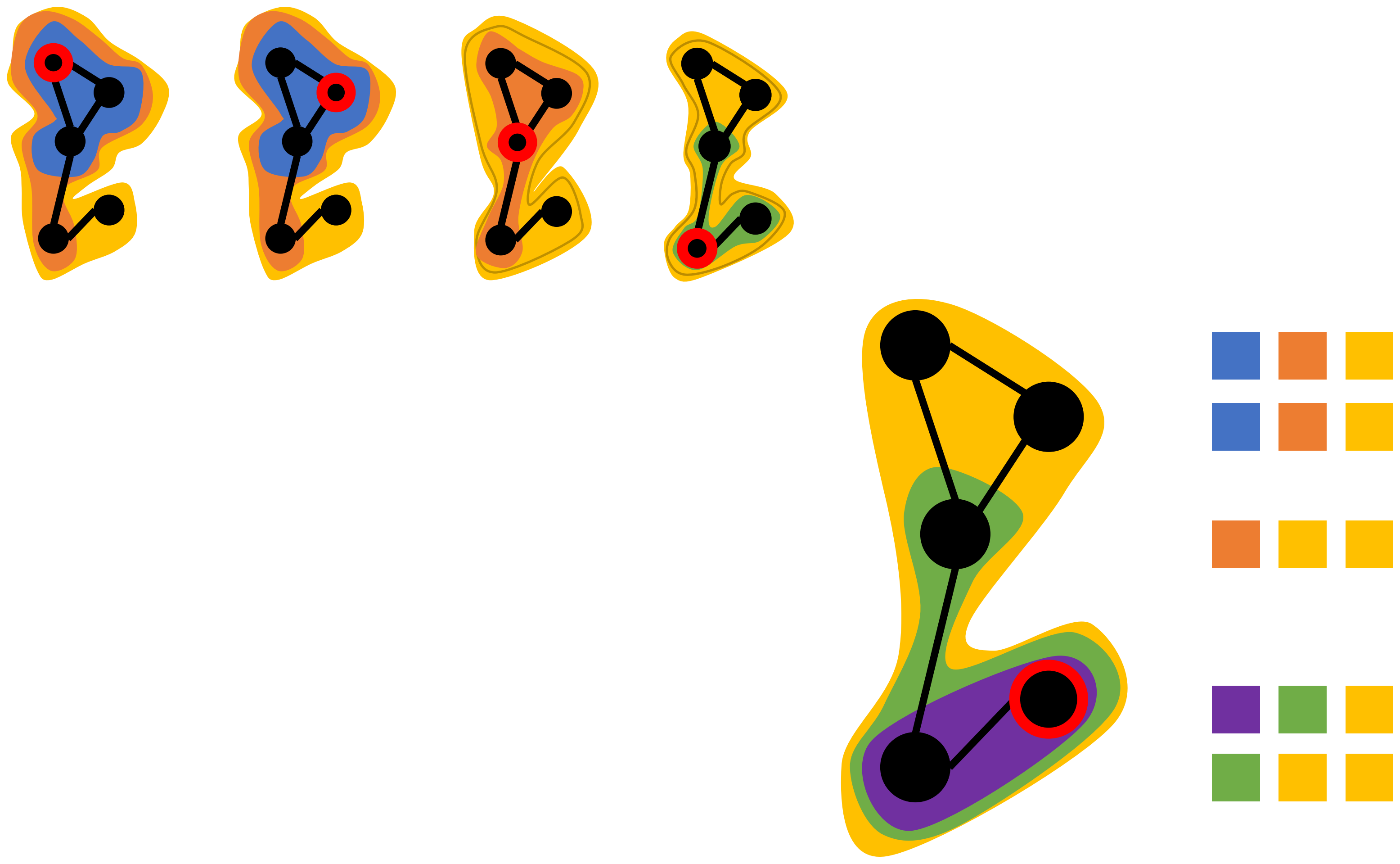


1.1 Weisfeiler-Lehman node embedding scheme
Idea

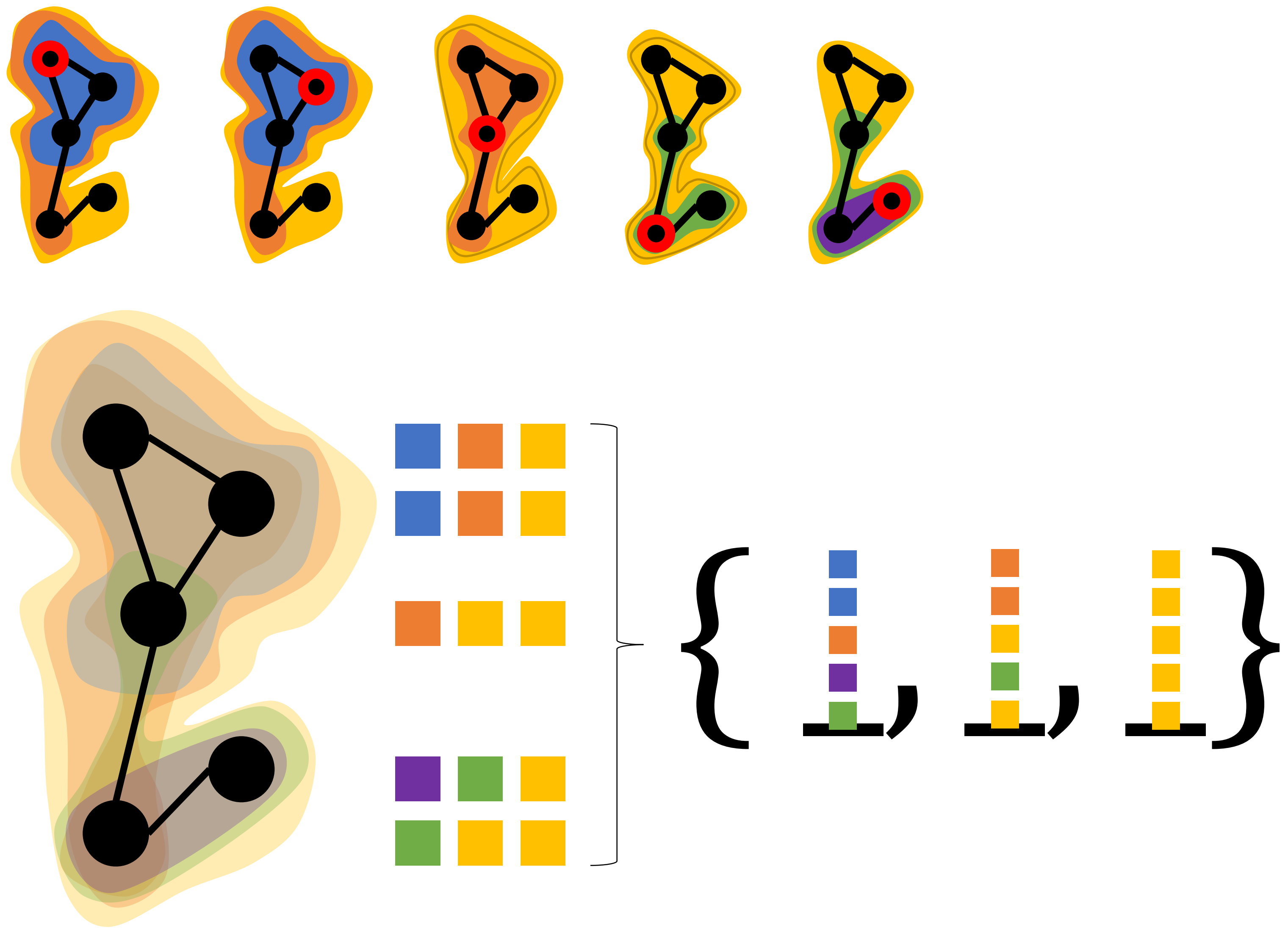




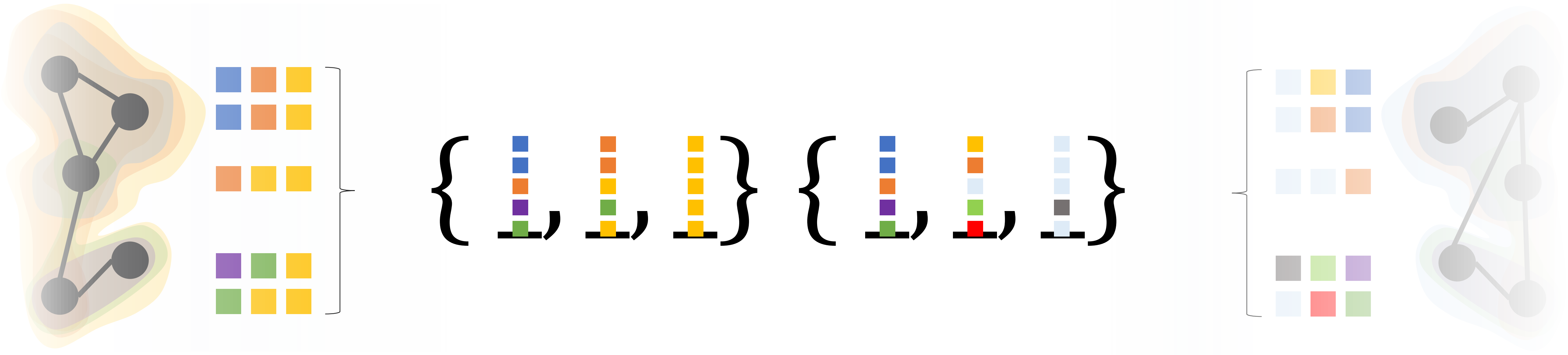
1.1 Weisfeiler-Lehman node embedding scheme
Idea



1.1 Weisfeiler-Lehman node embedding scheme
Idea

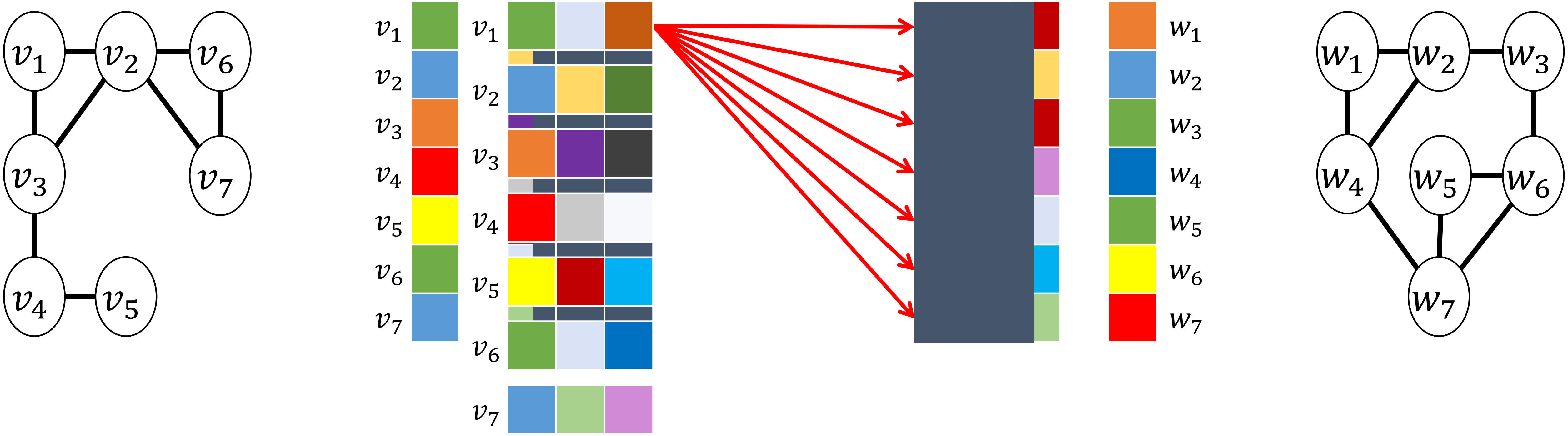


1.1 Weisfeiler-Lehman node embedding scheme
Idea

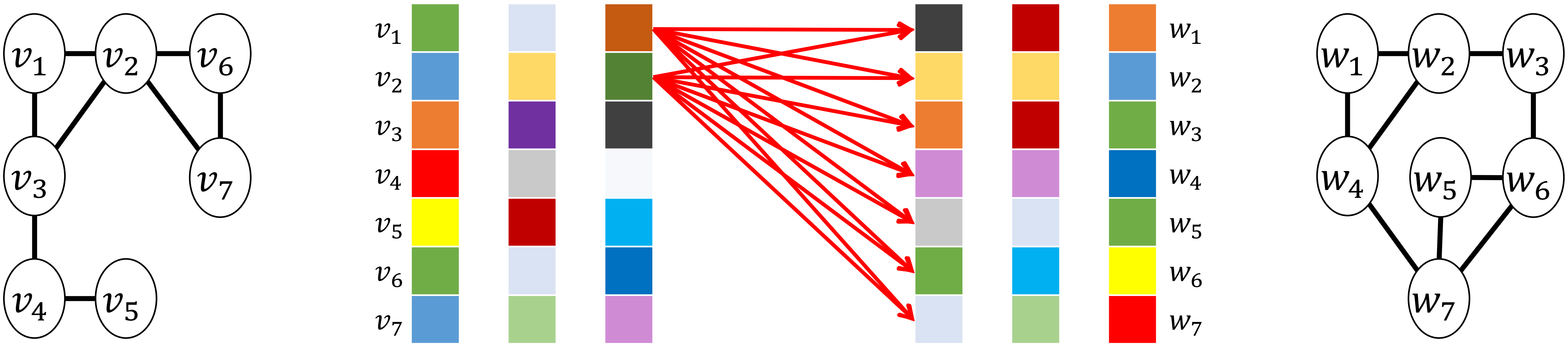


TODO: second version, encapsulating every step from
The method

1.1 Weisfeiler-Lehman node embedding scheme
Idea



1.1 Weisfeiler-Lehman node embedding scheme
Idea



1.1 Weisfeiler-Lehman node embedding scheme Theory

m: dimensionality of the node labels

n_G: cardinality of the node set

$N(v)$: neighbourhood of v

$$N^h(v) = \{x^h(u_1), \dots, x^h(u_{|N(v)|})\}$$

DISCRETE node labels	CONTINUOUS node labels
$x^{h+1}(v) = hash(x^h(v), N^h(v))$	$x^{h+1}(v) = x^h(v) + \sum_{u \in N(v)} x^h(u)$

1.1 Weisfeiler-Lehman node embedding scheme
Theory

m: dimensionality of the node labels

n_G: cardinality of the node set

$N(v)$: neighbourhood of v

$$N^h(v) = \left\{x^h(u_1), \dots, x^h\left(u_{|N(v)|}\right)\right\}$$

DISCRETE node labels	CONTINUOUS node labels
$x^{h+1}(v) = hash(x^h(v), N^h(v))$	$x^{h+1}(v) = x^h(v) + \sum_{u \in N(v)} w((u,v))x^h(u)$

1.1 Weisfeiler-Lehman node embedding scheme Theory

m: dimensionality of the node labels

n_G: cardinality of the vertex set

$N(v)$: neighbourhood of v

$$N^h(v) = \{l^h(u_0), \dots, l^h(u_{\deg(v)-1})\}$$

DISCRETE node labels	CONTINUOUS node labels
$l^{h+1}(v) = hash(l^h(v), N^h(v))$	$a^{h+1}(v) = \frac{1}{2}\left(a^h(v) + \frac{1}{\deg(v)} \sum_{u \in N(v)} w((u,v))\, a^h(u)\right) \in \mathbb{R}^m$

WL-features $X_G^h = \left[x^h(v_1), \dots, x^h(v_{n_G})\right]^T \in \mathbb{R}^{n_G \times m}$

Node embedding $f_{WL}^H : G \longrightarrow (X_G^0, \dots, X_G^H) \in \mathbb{R}^{n_G \times m(H+1)}$

2.2 Graph Wasserstein distance

$p \in [1, \infty)$: parameter for the $\boldsymbol{L^p}$ Wasserstein distance

$X \in \mathbb{R}^{n \times m}$

$X' \in \mathbb{R}^{n' \times m}$: set of vectors

$M = [d(x,y)]_{x \in X, y \in X'}$: distance matrix on a ground distance

$\Gamma(X, X')$: set of all transportaion plans

$\langle .,. \rangle$: Frobenius dot product

$$W_1(X, X') := \min_{P \in \Gamma(X, X')} \langle P, M \rangle$$

$$G, G' \quad \rightarrowtail \quad f_{WL}^H : G \to \mathbb{R}^{n_G \times m(H+1)} \quad \rightarrowtail \quad D_W^f(G, G') := W_1(f(G), f(G')) = \min_{P \in \Gamma(f(G), f(G'))} \langle P, M \rangle \quad \rightarrowtail \quad K_{WWL} = e^{-\lambda D_W^{f_{WL}}(G, G')}$$