Link Stream Compression for Multiscale Analysis of Temporal Interactions

Hindol Rakshit, Tiphaine Viard, and Robin Lamarche-Perrin

complexnetworks.fr

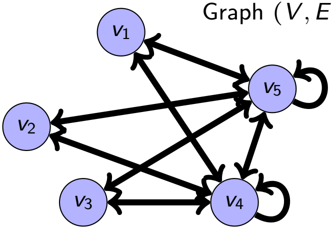
LIP6 UPMC SORBONNE UNIVERSITÉS CNRS

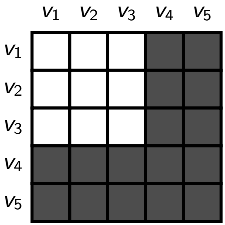
Vertices $V = \{v_1, \ldots, v_n\}$

Directed Edges $E \subseteq V \times V$

Vertex Neighbourhood $N(v) = (\{v' : (v, v') \in E\}, \{v' : (v', v) \in E\})$

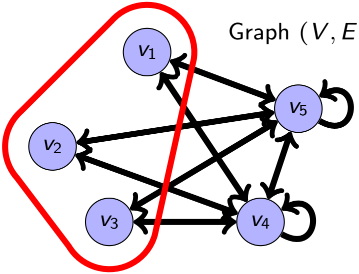
Graph (V, E)



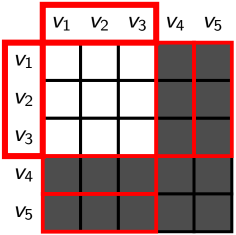


Lossless Modular Compression of Static Graphs

Vertices	$V = \{v_1, \dots, v_n\}$
Directed Edges	$E \subseteq V \times V$
Vertex Neighbourhood	$N(v) = (\{v' : (v, v') \in E\}, \{v' : (v', v) \in E\})$
Module	$M \subseteq V$ s.t. $\forall (v, v') \in M^2,\ N(v) = N(v')$

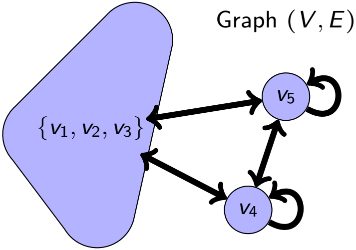


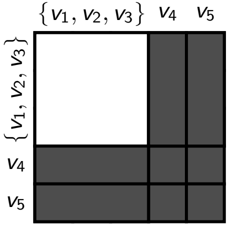
Graph (V, E)



Lossless Modular Compression of Static Graphs

Vertices	$V = \{v_1, \ldots, v_n\}$
Directed Edges	$E \subseteq V \times V$
Vertex Neighbourhood	$N(v) = (\{v' : (v, v') \in E\}, \{v' : (v', v) \in E\})$
Module	$M \subseteq V$ s.t. $\forall (v, v') \in M^2,\ N(v) = N(v')$





Lossless Modular Compression of Static Graphs

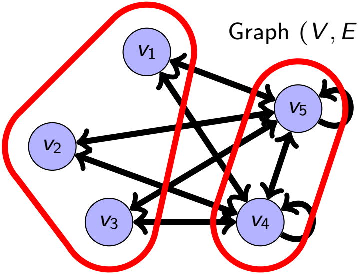
Vertices $V = \{v_1, \ldots, v_n\}$

Directed Edges $E \subseteq V \times V$

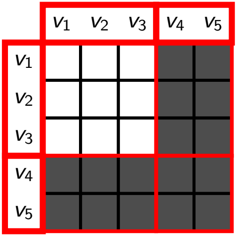
Vertex Neighbourhood $N(v) = (\{v' : (v, v') \in E\}, \{v' : (v', v) \in E\})$

Module $M \subseteq V$ s.t. $\forall (v, v') \in M^2,\ \ N(v) = N(v')$

Modular Partition $\mathcal{V} = \{M_1, \ldots, M_m\}$ s.t. $M_i \cap M_j = \varnothing$ and $\bigcup_i M_i = V$

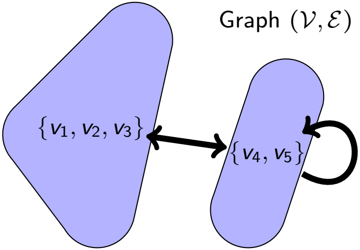


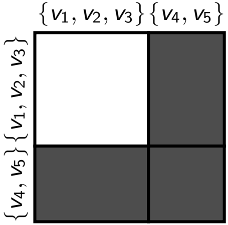
Graph (V, E)



Lossless Modular Compression of Static Graphs

Vertices	$V = \{v_1, \dots, v_n\}$
Directed Edges	$E \subseteq V \times V$
Vertex Neighbourhood	$N(v) = (\{v' : (v, v') \in E\}, \{v' : (v', v) \in E\})$
Module	$M \subseteq V$ s.t. $\forall (v, v') \in M^2,\ \ N(v) = N(v')$
Modular Partition	$\mathcal{V} = \{M_1, \dots, M_m\}$ s.t. $M_i \cap M_j = \varnothing$ and $\bigcup_i M_i = V$
Edge Compression	$\mathcal{E} = \{(M_i, M_j) \in \mathcal{V}^2 : \forall (v, v') \in M_i \times M_j, (v, v') \in E\}$





Lossless Modular Compression of Static Graphs

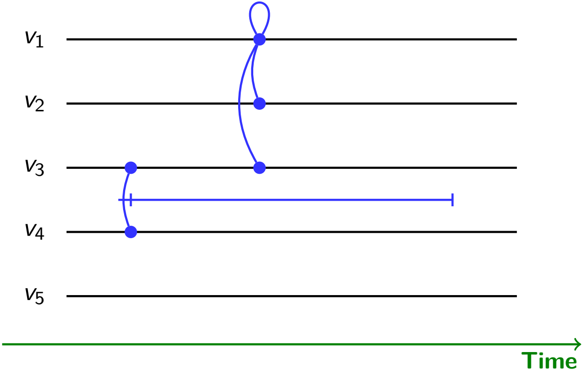
Vertices	$V = \{v_1, \ldots, v_n\}$
Directed Edges	$E \subseteq V \times V$
Vertex Neighbourhood	$N(v) = (\{v' : (v, v') \in E\}, \{v' : (v', v) \in E\})$
Module	$M \subseteq V$ s.t. $\forall (v, v') \in M^2,\ N(v) = N(v')$
Modular Partition	$\mathcal{V} = \{M_1, \ldots, M_m\}$ s.t. $M_i \cap M_j = \varnothing$ and $\bigcup_i M_i = V$
Edge Compression	$\mathcal{E} = \{(M_i, M_j) \in \mathcal{V}^2 : \forall (v, v') \in M_i \times M_j, (v, v') \in E\}$

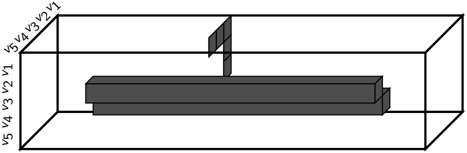
Combinatorial Problem

Given a graph (V, E), find a modular partition $\mathcal{V}^*$ with minimal size $|\mathcal{V}^*|$.

I. From Static to Dynamic Graphs

Compression of Dynamic Graphs





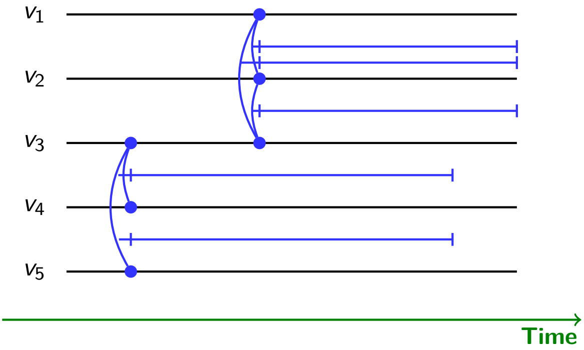
Vertices
$V = \{v_1, \ldots, v_n\}$

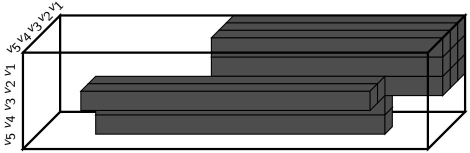
Time interval
$T = [\alpha, \omega] \in \mathbb{R}$

ST Edges
$E \subseteq V \times V \times T$

ST Neighbours
$$N(v,t) = (\{v' : (v, v', t) \in E\}, \{v' : (v', v, t) \in E\})$$

Compression of Dynamic Graphs





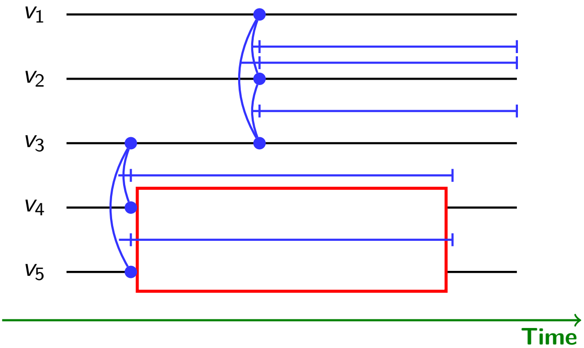
Vertices
$V = \{v_1, \ldots, v_n\}$

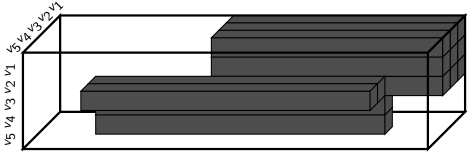
Time interval
$T = [\alpha, \omega] \in \mathbb{R}$

ST Edges
$E \subseteq V \times V \times T$

ST Neighbours
$$N(v,t) = (\{v' : (v, v', t) \in E\}, \{v' : (v', v, t) \in E\})$$

Compression of Dynamic Graphs





Vertices
$V = \{v_1, \dots, v_n\}$

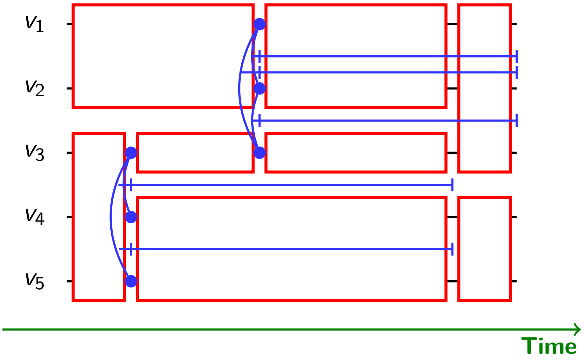
Time interval
$T = [\alpha, \omega] \in \mathbb{R}$

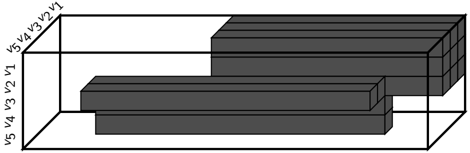
ST Edges
$E \subseteq V \times V \times T$

ST Neighbours
$$N(v,t) = (\{v' : (v, v', t) \in E\}, \{v' : (v', v, t) \in E\})$$

ST Module
$M \subseteq V, P = [\alpha', \omega'] \subseteq T$,
s.t. $\forall((v,t),(v',t')) \in (M \times P)^2,\ N(v,t) = N(v',t')$

Compression of Dynamic Graphs





Vertices
$V = \{v_1, \dots, v_n\}$

Time interval
$T = [\alpha, \omega] \in \mathbb{R}$

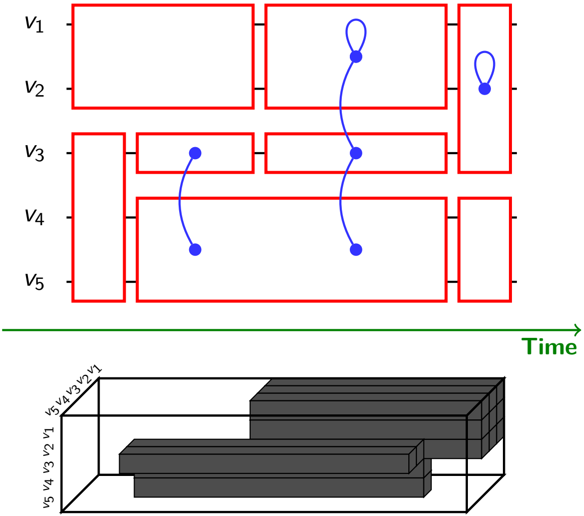
ST Edges
$E \subseteq V \times V \times T$

ST Neighbours
$N(v,t) = (\{v' : (v,v',t) \in E\},$
$\{v' : (v',v,t) \in E\})$

ST Module
$M \subseteq V, P = [\alpha', \omega'] \subseteq T,$
s.t. $\forall((v,t),(v',t')) \in$
$(M \times P)^2,\ N(v,t) = N(v',t')$

ST Modular Partition
$\mathcal{V} = \{(M_1, P_1), \dots, (M_m, P_m)\}$
s.t. $(M_i \times P_i) \cap (M_j \times P_j) = \emptyset$
and $\bigcup_i (M_i \times P_i) = (V \times T)$

Compression of Dynamic Graphs



Vertices
$V = \{v_1, \dots, v_n\}$

Time interval
$T = [\alpha, \omega] \in \mathbb{R}$

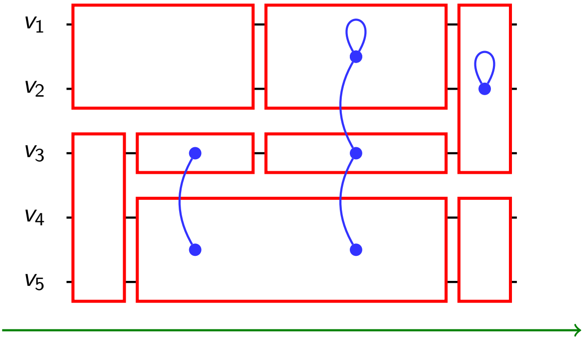
ST Edges
$E \subseteq V \times V \times T$

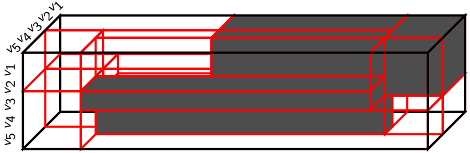
ST Neighbours
$N(v,t) = (\{v' : (v,v',t) \in E\},$
$\{v' : (v',v,t) \in E\})$

ST Module
$M \subseteq V, P = [\alpha', \omega'] \subseteq T,$
s.t. $\forall((v,t),(v',t')) \in$
$(M \times P)^2,\ N(v,t) = N(v',t')$

ST Modular Partition
$\mathcal{V} = \{(M_1, P_1), \dots, (M_m, P_m)\}$
s.t. $(M_i \times P_i) \cap (M_j \times P_j) = \varnothing$
and $\bigcup_i (M_i \times P_i) = (V \times T)$

ST Edge Compression
$\mathcal{E} = \{((M_i, P_i), (M_j, P_j)) \in \mathcal{V}^2 :$
$\forall (v, v', t) \in M_i \times M_j \times P_i \cap P_j,$
$(v, v', t) \in E\}$





Vertices
$V = \{v_1, \ldots, v_n\}$

Time interval
$T = [\alpha, \omega] \in \mathbb{R}$

ST Edges
$E \subseteq V \times V \times T$

ST Neighbours
$N(v,t) = (\{v' : (v, v', t) \in E\},$
$\{v' : (v', v, t) \in E\})$

ST Module
$M \subseteq V, P = [\alpha', \omega'] \subseteq T,$
s.t. $\forall((v,t),(v',t')) \in$
$(M \times P)^2,\ N(v,t) = N(v',t')$

ST Modular Partition
$\mathcal{V} = \{(M_1, P_1), \ldots, (M_m, P_m)\}$
s.t. $(M_i \times P_i) \cap (M_j \times P_j) = \varnothing$
and $\bigcup_i (M_i \times P_i) = (V \times T)$

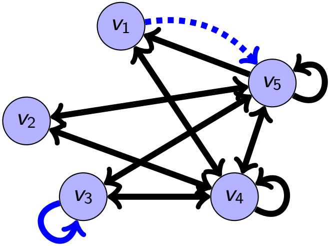
ST Edge Compression
$\mathcal{E} = \{((M_i, P_i), (M_j, P_j)) \in \mathcal{V}^2 :$
$\forall(v, v', t) \in M_i \times M_j \times P_i \cap P_j,$
$(v, v', t) \in E\}$

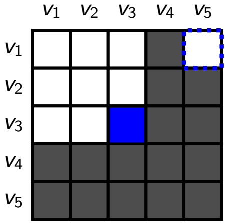
From Static to Dynamic Notations

	Static Graphs	**Dynamic Graphs**
Edges	$E \subseteq V \times V$	$E \subseteq V \times V \times T$
Neighbourhood	$N(v) = (\{v' : (v, v') \in E\}, \{v' : (v', v) \in E\})$	$N(v, t) = (\{v' : (v, v', t) \in E\}, \{v' : (v', v, t) \in E\})$
Module	$M \subseteq V$, s.t. $\forall (v, v') \in M^2$, $N(v) = N(v')$	$M \subseteq V, P = [\alpha', \omega'] \subseteq T$, s.t. $\forall ((v, t), (v', t')) \in (M \times P)^2$, $N(v, t) = N(v', t')$
Modular Partition	$\mathcal{V} = \{M_1, \ldots, M_m\}$, s.t. $M_i \cap M_j = \varnothing$ and $\bigcup_i M_i = V$	$\mathcal{V} = \{(M_1, P_1), \ldots, (M_m, P_m)\}$, s.t. $(M_i \times P_i) \cap (M_j \times P_j) = \varnothing$ and $\bigcup_i (M_i \times P_i) = (V \times T)$
Edge Compression	$\mathcal{E} = \{(M_i, M_j) \in \mathcal{V}^2 : \forall (v, v') \in M_i \times M_j, (v, v') \in E\}$	$\mathcal{E} = \{((M_i, P_i), (M_j, P_j)) \in \mathcal{V}^2 : \forall (v, v', t) \in M_i \times M_j \times P_i \cap P_j, (v, v', t) \in E\}$

II. From Lossless to Lossy Compression

Lossy Compression of Static Graphs

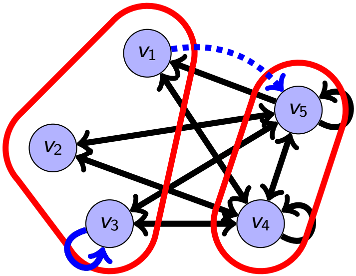


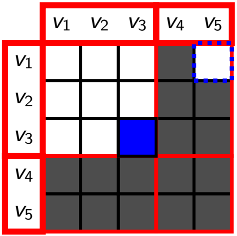


Lossy Compression of Static Graphs

Aggregate $M \subseteq V$

Partition $\mathcal{V} = \{M_1, \ldots, M_m\}$ s.t. $M_i \cap M_j = \varnothing$ and $\bigcup_i M_i = V$



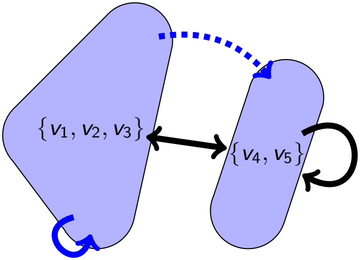


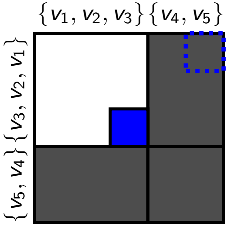
Lossy Compression of Static Graphs

Aggregate $M \subseteq V$

Partition $\mathcal{V} = \{M_1, \ldots, M_m\}$ s.t. $M_i \cap M_j = \varnothing$ and $\bigcup_i M_i = V$

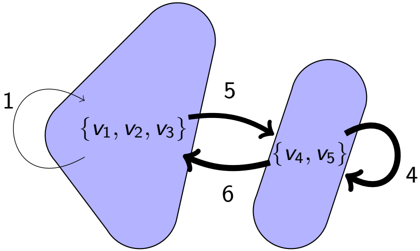
Edge Compression $\mathcal{E} = \{(M_i, M_j) \in \mathcal{V}^2 : \delta(M_i, M_j) > \frac{1}{2}\}$

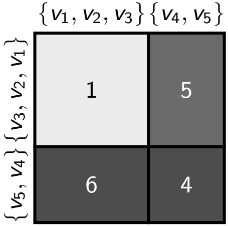




Lossy Compression of Static Graphs

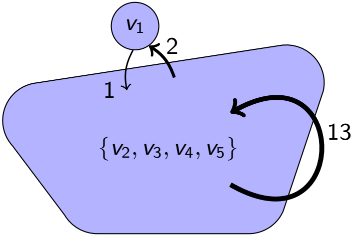
Aggregate	$M \subseteq V$
Partition	$\mathcal{V} = \{M_1, \dots, M_m\}$ s.t. $M_i \cap M_j = \varnothing$ and $\bigcup_i M_i = V$
Edge Compression	$\mathcal{E} = \{(M_i, M_j) \in \mathcal{V}^2 : w(M_i, M_j) > 0\}$
Aggregate Weight	$w(M_i, M_j) = \sum_{(v,v') \in M_i \times M_j} w(v, v')$
Aggregate Density	$\delta(M_i, M_j) = \frac{w(M_i, M_j)}{\lvert M_i \rvert \, \lvert M_j \rvert}$





Lossy Compression of Static Graphs

Aggregate	$M \subseteq V$
Partition	$\mathcal{V} = \{M_1, \ldots, M_m\}$ s.t. $M_i \cap M_j = \varnothing$ and $\bigcup_i M_i = V$
Edge Compression	$\mathcal{E} = \{(M_i, M_j) \in \mathcal{V}^2 : w(M_i, M_j) > 0\}$
Aggregate Weight	$w(M_i, M_j) = \sum_{(v,v') \in M_i \times M_j} w(v, v')$
Aggregate Density	$\delta(M_i, M_j) = \frac{w(M_i, M_j)}{\lvert M_i \rvert \, \lvert M_j \rvert}$



	$\{v_1\}$	$\{v_2, v_3, v_4, v_5\}$
$\{v_1\}$	0	1
$\{v_2, v_3, v_4, v_5\}$	2	13

Lossy Compression of Static Graphs

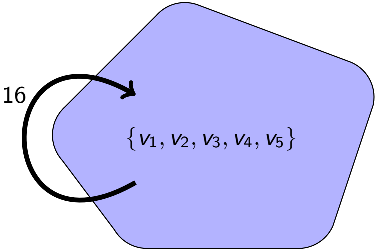
Aggregate $M \subseteq V$

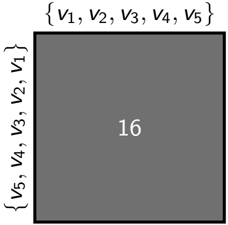
Partition $\mathcal{V} = \{M_1, \ldots, M_m\}$ s.t. $M_i \cap M_j = \emptyset$ and $\bigcup_i M_i = V$

Edge Compression $\mathcal{E} = \{(M_i, M_j) \in \mathcal{V}^2 : w(M_i, M_j) > 0\}$

Aggregate Weight $w(M_i, M_j) = \sum_{(v,v') \in M_i \times M_j} w(v, v')$

Aggregate Density $\delta(M_i, M_j) = \frac{w(M_i, M_j)}{|M_i|\,|M_j|}$





Lossy Compression of Static Graphs

Aggregate $M \subseteq V$

Partition $\mathcal{V} = \{M_1, \ldots, M_m\}$ s.t. $M_i \cap M_j = \emptyset$ and $\bigcup_i M_i = V$

Edge Compression $\mathcal{E} = \{(M_i, M_j) \in \mathcal{V}^2 : w(M_i, M_j) > 0\}$

Aggregate Weight $w(M_i, M_j) = \sum_{(v,v') \in M_i \times M_j} w(v, v')$

Aggregate Density $\delta(M_i, M_j) = \frac{w(M_i, M_j)}{|M_i|\,|M_j|}$

Combinatorial Problem
Given a graph (V, E), an information measure f, and a threshold $\tau \in \mathbb{R}^+$, find a partition $\mathcal{V}^*$ such that $f(\mathcal{V}^*) \leq \tau$ with minimal size $|\mathcal{V}^*|$.

Cosine Similarity on neighbours vectors

Pearson Coefficient on common neighbours

Jaccard Similarity Index on neighbourhoods

$$Q_J(M) = \min_{(v,v')\in M^2} \frac{|N(v) \cap N(v')|}{|N(v) \cup N(v')|} \in [0,1]$$

Average Squared Errors of densities

$$E_\delta(M_i, M_j) = \frac{1}{|M_i|\,|M_j|} \sum_{\substack{v\in M_i \\ v'\in M_j}} \left(\delta(v,v') - \delta(M_i, M_j)\right)^2$$

Log-likelihood of the Stochastic Block Model

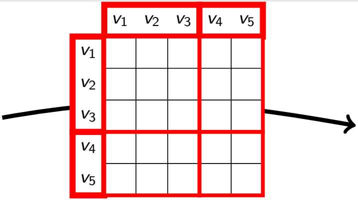
$$\begin{aligned}\log \mathcal{L}(\mathcal{V}) = \sum_{(M_i,M_j)\in\mathcal{V}} \quad & w(M_i,M_j)\log w(M_i,M_j) \\ & +(|M_i|\,|M_j| - w(M_i,M_j))\log(|M_i|\,|M_j| - w(M_i,M_j)) \\ & -|M_i|\,|M_j|\log|M_i|\,|M_j|\end{aligned}$$

$$\begin{array}{c|ccccc} & v_1 & v_2 & v_3 & v_4 & v_5 \\ \hline v_1 & 0 & 0 & 0 & 1 & 0 \\ v_2 & 0 & 0 & 0 & 1 & 1 \\ v_3 & 0 & 0 & 1 & 1 & 1 \\ v_4 & 1 & 1 & 1 & 1 & 1 \\ v_5 & 1 & 1 & 1 & 1 & 1 \end{array}$$

Empirical Distribution: $X \in V^2$

$$p_X(v, v') = \frac{w(v, v')}{|E|}$$

	v_1	v_2	v_3	v_4	v_5
v_1	0	0	0	1	0
v_2	0	0	0	1	1
v_3	0	0	1	1	1
v_4	1	1	1	1	1
v_5	1	1	1	1	1

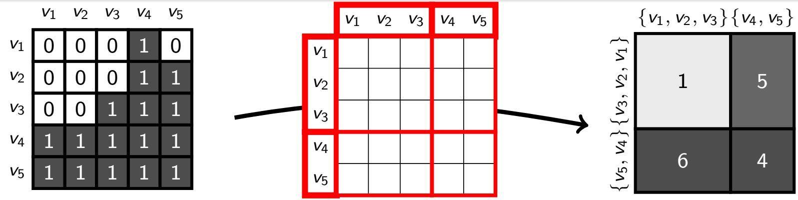


Empirical Distribution: $X \in V^2$

$$p_X(v, v') = \frac{w(v, v')}{|E|}$$

Compression Variable: $\hat{X} \in \mathcal{V}^2$

$$p_{\hat{X}|X}(M_i, M_j | v, v') = \mathbf{1}_{M_i \times M_j}(v, v')$$



Empirical Distribution: $X \in V^2$

$$p_X(v, v') = \frac{w(v, v')}{|E|}$$

Compression Variable: $\hat{X} \in \mathcal{V}^2$

$$p_{\hat{X}|X}(M_i, M_j | v, v') = \mathbf{1}_{M_i \times M_j}(v, v')$$

Compressed Distribution

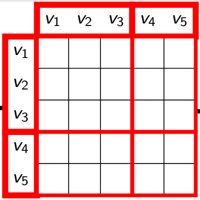
$$p_{\hat{X}}(M_i, M_j) = \frac{w(M_i, M_j)}{|E|}$$

Quantifying the Quality/Error of Partitions (using Information Theory)

	v_1	v_2	v_3	v_4	v_5
v_1	0	0	0	1	0
v_2	0	0	0	1	1
v_3	0	0	1	1	1
v_4	1	1	1	1	1
v_5	1	1	1	1	1

Empirical Distribution: $X \in V^2$

$$p_X(v, v') = \frac{w(v, v')}{|E|}$$



Compression Variable: $\hat{X} \in \mathcal{V}^2$

$$p_{\hat{X}|X}(M_i, M_j | v, v') = \mathbf{1}_{M_i \times M_j}(v, v')$$

	$\{v_1, v_2, v_3\}$	$\{v_4, v_5\}$
$\{v_3, v_2, v_1\}$	1	5
$\{v_5, v_4\}$	6	4

Compressed Distribution

$$p_{\hat{X}}(M_i, M_j) = \frac{w(M_i, M_j)}{|E|}$$

	v_1	v_2	v_3	v_4	v_5
v_1	1/20	1/20	1/20	1/20	1/20
v_2	1/20	1/20	1/20	1/20	1/20
v_3	1/20	1/20	1/20	1/20	1/20
v_4	1/20	1/20	1/20	1/20	1/20
v_5	1/20	1/20	1/20	1/20	1/20

Decompression Variable: $X^\star \in V^2$

$$u_{X^\star}(v, v') = \frac{1}{|V|^2}$$

	v_1	v_2	v_3	v_4	v_5
v_1	0	0	0	1	0
v_2	0	0	0	1	1
v_3	0	0	1	1	1
v_4	1	1	1	1	1
v_5	1	1	1	1	1

Empirical Distribution: $X \in V^2$

$$p_X(v, v') = \frac{w(v, v')}{|E|}$$

	v_1	v_2	v_3	v_4	v_5
v_1					
v_2					
v_3					
v_4					
v_5					

Compression Variable: $\hat{X} \in \mathcal{V}^2$

$$p_{\hat{X}|X}(M_i, M_j | v, v') = \mathbf{1}_{M_i \times M_j}(v, v')$$

	$\{v_1, v_2, v_3\}$	$\{v_4, v_5\}$
$\{v_1, v_2, v_3\}$	1	5
$\{v_4, v_5\}$	6	4

Compressed Distribution

$$p_{\hat{X}}(M_i, M_j) = \frac{w(M_i, M_j)}{|E|}$$

	v_1	v_2	v_3	v_4	v_5
v_1	1/20	1/20	1/20	1/20	1/20
v_2	1/20	1/20	1/20	1/20	1/20
v_3	1/20	1/20	1/20	1/20	1/20
v_4	1/20	1/20	1/20	1/20	1/20
v_5	1/20	1/20	1/20	1/20	1/20

Decompression Variable: $X^\star \in V^2$

$$u_{X^\star}(v, v') = \frac{1}{|V|^2}$$

	v_1	v_2	v_3	v_4	v_5
v_1	1/9	1/9	1/9	5/6	5/6
v_2	1/9	1/9	1/9	5/6	5/6
v_3	1/9	1/9	1/9	5/6	5/6
v_4	1	1	1	1	1
v_5	1	1	1	1	1

Decompressed Distribution

$$q_X(v, v') = \frac{w(M_i, M_j)}{|M_i|\,|M_j|\,|E|}$$

Quantifying the Quality/Error of Partitions (using Information Theory)

	v_1	v_2	v_3	v_4	v_5
v_1	0	0	0	1	0
v_2	0	0	0	1	1
v_3	0	0	1	1	1
v_4	1	1	1	1	1
v_5	1	1	1	1	1

Empirical Distribution: $X \in V^2$

$$p_X(v, v') = \frac{w(v, v')}{|E|}$$

	v_1	v_2	v_3	v_4	v_5
v_1					
v_2					
v_3					
v_4					
v_5					

Compression Variable: $\hat{X} \in \mathcal{V}^2$

$$p_{\hat{X}|X}(M_i, M_j | v, v') = \mathbf{1}_{M_i \times M_j}(v, v')$$

	$\{v_1, v_2, v_3\}$	$\{v_4, v_5\}$
$\{v_3, v_2, v_1\}$	1	5
$\{v_5, v_4\}$	6	4

Compressed Distribution

$$p_{\hat{X}}(M_i, M_j) = \frac{w(M_i, M_j)}{|E|}$$

Information Loss
$D_{KL}(p_X \| q_X)$

	v_1	v_2	v_3	v_4	v_5
v_1	$1/9$	$1/9$	$1/9$	$5/6$	$5/6$
v_2	$1/9$	$1/9$	$1/9$	$5/6$	$5/6$
v_3	$1/9$	$1/9$	$1/9$	$5/6$	$5/6$
v_4	1	1	1	1	1
v_5	1	1	1	1	1

Decompressed Distribution

$$q_X(v, v') = \frac{w(M_i, M_j)}{|M_i|\,|M_j|\,|E|}$$

	v_1	v_2	v_3	v_4	v_5
v_1	$1/20$	$1/20$	$1/20$	$1/20$	$1/20$
v_2	$1/20$	$1/20$	$1/20$	$1/20$	$1/20$
v_3	$1/20$	$1/20$	$1/20$	$1/20$	$1/20$
v_4	$1/20$	$1/20$	$1/20$	$1/20$	$1/20$
v_5	$1/20$	$1/20$	$1/20$	$1/20$	$1/20$

Decompression Variable: $X^\star \in V^2$

$$u_{X^\star}(v, v') = \frac{1}{|V|^2}$$

	v_1	v_2	v_3	v_4	v_5
v_1	0	0	0	1	0
v_2	0	0	0	1	1
v_3	0	0	1	1	1
v_4	1	1	1	1	1
v_5	1	1	1	1	1

	v_1	v_2	v_3	v_4	v_5
v_1					
v_2					
v_3					
v_4					
v_5					

	$\{v_1, v_2, v_3\}$	$\{v_4, v_5\}$
$\{v_1, v_2, v_3\}$	1	5
$\{v_4, v_5\}$	6	4

Emp

istribution

$$w(M_i, M_j) \over |E|$$

Informat

$D_{KL}(p_X \|$

Information Loss using Kullback-Leibler Divergence

$$D_{KL}(p_X \| q_X) = \frac{1}{|E|} \sum_{\substack{(M_i, M_j) \in \mathcal{V}^2 \\ (v, v') \in M_i \times M_j}} w(v, v') \log_2 \left(\frac{w(v, v')}{w(M_i, M_j)} |M_i| \, |M_j| \right)$$

	v_1	v_2	v_3	v_4	v_5
v_2	1/9	1/9	1/9	5/6	5/6
v_3	1/9	1/9	1/9	5/6	5/6
v_4	1	1	1	1	1
v_5	1	1	1	1	1

	v_1	v_2	v_3	v_4	v_5
v_2	1/20	1/20	1/20	1/20	1/20
v_3	1/20	1/20	1/20	1/20	1/20
v_4	1/20	1/20	1/20	1/20	1/20
v_5	1/20	1/20	1/20	1/20	1/20

Decompressed Distribution

$$q_X(v, v') = \frac{w(M_i, M_j)}{|M_i| \, |M_j| \, |E|}$$

Decompression Variable: $X^\star \in V^2$

$$u_{X^\star}(v, v') = \frac{1}{|V|^2}$$

Quantifying the Quality/Error of Partitions (using Information Theory)

	v_1	v_2	v_3	v_4	v_5
v_1	1	1	3	18	6
v_2	0	2	0	19	11
v_3	1	0	9	19	11
v_4	8	9	10	21	19
v_5	11	12	10	20	20

Empirical Distribution: $X \in V^2$

$$p_X(v, v') = \frac{w(v, v')}{|E|}$$

	v_1	v_2	v_3	v_4	v_5
v_1					
v_2					
v_3					
v_4					
v_5					

Compression Variable: $\hat{X} \in \mathcal{V}^2$

$$p_{\hat{X}|X}(M_i, M_j | v, v') = \mathbf{1}_{M_i \times M_j}(v, v')$$

	$\{v_1, v_2, v_3\}$	$\{v_4, v_5\}$
$\{v_1, v_2, v_3\}$	18	84
$\{v_4, v_5\}$	60	80

Compressed Distribution

$$p_{\hat{X}}(M_i, M_j) = \frac{w(M_i, M_j)}{|E|}$$

Information Loss
$D_{KL}(p_X \| q_X)$

	v_1	v_2	v_3	v_4	v_5
v_1	2	2	2	14	14
v_2	2	2	2	14	14
v_3	2	2	2	14	14
v_4	10	10	10	20	20
v_5	10	10	10	20	20

Decompressed Distribution

$$q_X(v, v') = \frac{w(M_i, M_j)}{|M_i|\,|M_j|\,|E|}$$

	v_1	v_2	v_3	v_4	v_5
v_1	1/20	1/20	1/20	1/20	1/20
v_2	1/20	1/20	1/20	1/20	1/20
v_3	1/20	1/20	1/20	1/20	1/20
v_4	1/20	1/20	1/20	1/20	1/20
v_5	1/20	1/20	1/20	1/20	1/20

Decompression Variable: $X^\star \in V^2$

$$u_{X^\star}(v, v') = \frac{1}{|V|^2}$$

Quantifying the Quality/Error of Partitions (using Information Theory)

	v_1	v_2	v_3	v_4	v_5
v_1	1	1	3	18	6
v_2	0	2	0	19	11
v_3	1	0	9	19	11
v_4	8	9	10	21	19
v_5	11	12	10	20	20

Empirical Distribution: $X \in V^2$

$$p_X(v, v') = \frac{w(v, v')}{|E|}$$

	v_1	v_2	v_3	v_4	v_5
v_1					
v_2					
v_3					
v_4					
v_5					

Compression Variable: $\hat{X} \in \mathcal{V}^2$

$$p_{\hat{X}|X}(M_i, M_j | v, v') = \mathbf{1}_{M_i \times M_j}(v, v')$$

	$\{v_1, v_2, v_3\}$	$\{v_4, v_5\}$
$\{v_1, v_2, v_3\}$	18	84
$\{v_4, v_5\}$	60	80

Compressed Distribution

$$p_{\hat{X}}(M_i, M_j) = \frac{w(M_i, M_j)}{|E|}$$

Information Loss
$D_{KL}(p_X \| q_X)$

	v_1	v_2	v_3	v_4	v_5	
v_1						29
v_2						32
v_3						41
v_4						67
v_5						73
	21	33	24	97	67	

External Information

	v_1	v_2	v_3	v_4	v_5
v_1	3	4	3	12	8
v_2	3	4	3	13	9
v_3	4	6	4	17	11
v_4	6	9	7	27	19
v_5	6	10	7	29	20

Decompression Variable: $X^\star \in V^2$

$$u_{X^\star}(v, v') = \frac{w(v, .)\, w(., v')}{|E|}$$

	v_1	v_2	v_3	v_4	v_5
v_1	1	2	2	14	10
v_2	2	2	2	16	11
v_3	2	3	2	20	14
v_4	8	12	9	23	16
v_5	8	13	10	25	17

Decompressed Distribution

$$q_X(v, v') = \frac{w(M_i, M_j)\, w(v, .)\, w(., v')}{w(M_i, .)\, w(., M_j)\, |E|}$$

Lossy Compression of Dynamic Graphs

Information Loss in the Static case

$$D_{KL}(p_X \| q_X) = \frac{1}{|E|} \sum_{\substack{M_i \in \mathcal{V} \\ M_j \in \mathcal{V} \\ (v,v') \in M_i \times M_j}} w(v,v') \log_2 \left(\frac{w(v,v')}{w(M_i,M_j)} |M_i||M_j| \right)$$

where

$$w(v,v') = \mathbf{1}_E(v,v')$$

$$w(M_i,M_j) = \sum_{(v,v') \in M_i \times M_j} w(v,v')$$

Information Loss in the Dynamic case

$$D_{KL}(p_X \| q_X) = \frac{1}{|E|} \sum_{\substack{(M_i,P_i) \in \mathcal{V} \\ (M_j,P_j) \in \mathcal{V} \\ (v,v') \in M_i \times M_j}} \int_{t \in P_i \cap P_j} w(v,v',t) \log_2 \left(\frac{w(v,v',t)}{w(M_i,M_j,P_i \cap P_j)} |M_i||M_j||P_i \cap P_j| \right) dt$$

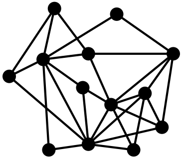
where

$$w(v,v',t) = \mathbf{1}_E(v,v',t)$$

$$w(M_i,M_j,P_i \cap P_j) = \sum_{(v,v') \in M_i \times M_j} \int_{t \in P_i \cap P_j} w(v,v',t)\, dt$$

III. From Modular to Power Graph Compression

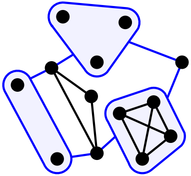
Constraints in Graph Compression



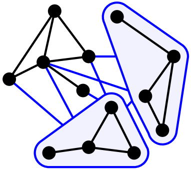
Initial graph

01000111 01110010
01100001 01110000
01101000 01100101
00100000 01111010
01101001 01110000
01110000 11000011
10101001 00100001

Unconstrained compression
(no constraint)

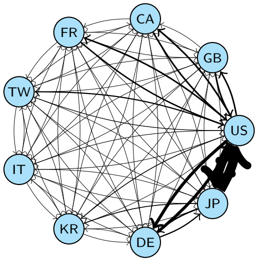


Modular Decomposition
(generic constraints)



Graph Rewriting
(strong constraints)

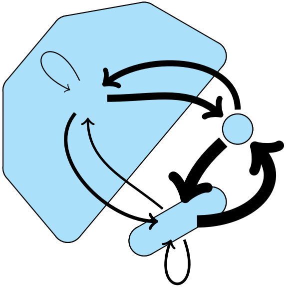
Power Graph Decomposition



	GB	CA	FR	TW	IT	KR	DE	JP	US
GB		3	5	1	2	0	11	23	82
CA	3		3	2	1	0	6	15	89
FR	5	3		1	3	1	14	28	83
TW	2	3	2		1	3	4	22	62
IT	2	1	3	1		0	7	12	31
KR	2	1	2	2	1		3	47	44
DE	11	6	12	2	6	1		78	167
JP	24	14	23	9	9	14	66		504
US	86	87	75	37	29	16	161	519	

National Patent Citations. Unit: 100 patents; Period: 1990–1999;
Source: NBER U.S. Patent Citations Data File (http://www.nber.org/patents/)

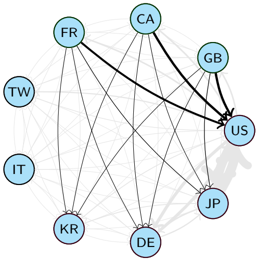
Power Graph Decomposition



	GB CA FR TW IT KR	DE JP	US
GB CA FR TW IT KR	59	192	391
DE JP	131	144	671
US	330	680	

National Patent Citations. Unit: 100 patents; Period: 1990–1999;
Source: NBER U.S. Patent Citations Data File (http://www.nber.org/patents/)

Aggregate $M \subseteq V, N \subseteq V$



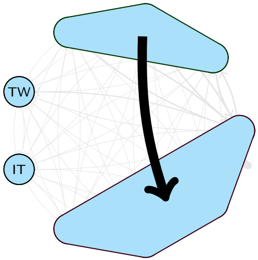
	GB	CA	FR	TW	IT	KR	DE	JP	US
GB		3	5	1	2	0	11	23	82
CA	3		3	2	1	0	6	15	89
FR	5	3		1	3	1	14	28	83
TW	2	3	2		1	3	4	22	62
IT	2	1	3	1		0	7	12	31
KR	2	1	2	2	1		3	47	44
DE	11	6	12	2	6	1		78	167
JP	24	14	23	9	9	14	66		504
US	86	87	75	37	29	16	161	519	

National Patent Citations. Unit: 100 patents; Period: 1990–1999;
Source: NBER U.S. Patent Citations Data File (http://www.nber.org/patents/)

Power Graph Decomposition

Aggregate $M \subseteq V, N \subseteq V$

Aggregate Weight $w(M_i, N_i) = \sum_{(v,v') \in M_i \times N_i} w(v, v')$



<table>
<tr><th></th><th>GB</th><th>CA</th><th>FR</th><th>TW</th><th>IT</th><th>KR</th><th>DE</th><th>JP</th><th>US</th></tr>
<tr><th>GB</th><td></td><td>3</td><td>5</td><td>1</td><td>2</td><td colspan="4" rowspan="3">352</td></tr>
<tr><th>CA</th><td>3</td><td></td><td>3</td><td>2</td><td>1</td></tr>
<tr><th>FR</th><td>5</td><td>3</td><td></td><td>1</td><td>3</td></tr>
<tr><th>TW</th><td>2</td><td>3</td><td>2</td><td></td><td>1</td><td>3</td><td>4</td><td>22</td><td>62</td></tr>
<tr><th>IT</th><td>2</td><td>1</td><td>3</td><td>1</td><td></td><td>0</td><td>7</td><td>12</td><td>31</td></tr>
<tr><th>KR</th><td>2</td><td>1</td><td>2</td><td>2</td><td>1</td><td></td><td>3</td><td>47</td><td>44</td></tr>
<tr><th>DE</th><td>11</td><td>6</td><td>12</td><td>2</td><td>6</td><td>1</td><td></td><td>78</td><td>167</td></tr>
<tr><th>JP</th><td>24</td><td>14</td><td>23</td><td>9</td><td>9</td><td>14</td><td>66</td><td></td><td>504</td></tr>
<tr><th>US</th><td>86</td><td>87</td><td>75</td><td>37</td><td>29</td><td>16</td><td>161</td><td>519</td><td></td></tr>
</table>

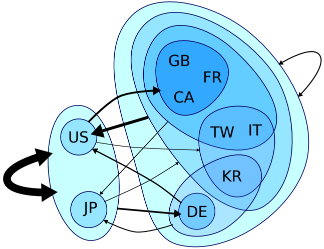
National Patent Citations. Unit: 100 patents; Period: 1990–1999;
Source: NBER U.S. Patent Citations Data File (http://www.nber.org/patents/)

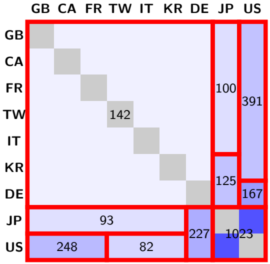
Power Graph Decomposition

Aggregate $M \subseteq V, N \subseteq V$

Aggregate Weight $w(M_i, N_i) = \sum_{(v,v') \in M_i \times N_i} w(v, v')$

Partition $\mathcal{V} = \{(M_1, N_1), \ldots, (M_m, N_m)\}$
s.t. $(M_i \times N_i) \cap (M_j \times N_j) = \varnothing$
and $\bigcup_i (M_i \times N_i) = V^2$





National Patent Citations. Unit: 100 patents; Period: 1990–1999;
Source: NBER U.S. Patent Citations Data File (http://www.nber.org/patents/)

Link Stream Compression for Multiscale Analysis of Temporal Interactions

Hindol Rakshit, Tiphaine Viard, and Robin Lamarche-Perrin

complexnetworks.fr