Gaussian Covariance Decomposition for PC-algorithm

Dhafer Malouche

LEGI-EPT-ESSAI
Uty, November 7th Carthage, Tunisia,

Mathematical Aspects of Graphical Models

Durham, July 2008

Graphical Models : Notations

- $G = (V, E)$ an undirected graph
  - V set of *vertices*
  - E set of *edges* : $(u, v) \in E \iff (v, u) \in E$
- $u \sim_G v \iff (u, v) \in E$,
  $\mathrm{ne}_G(u) = \{v,\ v \sim_G u\}$ *neighbors* of u.
- $p(u, v, G) = (u_0, u_1, \ldots, u_n)$ is a *path* between u and v in G if $u_0 = u$, $u_n = v$ and $\forall\, i = 0, \ldots (n-1)$, $u_i \sim_G u_{i+1}$, and the u_i's are distinct.
  $|p(u, v, G)| = n$ is the *length* of $p(u, v, G)$.
- $G = (V, E)$ and $G' = (V, E')$, $G \subseteq G' \iff E \subseteq E'$

Graphical Models : Markov Property (MP)

- $\mathbf{X} = \mathbf{X}_V = (X_u,\ u \in V)' \sim P = \mathcal{N}_{|V|}(\mu, \Sigma)$ i s Gaussian r.v. $\Sigma : |V| \times |V|$ covariance matrix and $K = \Sigma^{-1}$ is the precision matrix.
- $K \longmapsto G(K) = G$:

$$u \not\sim_G v \iff k_{uv} = 0 \quad (\underline{Pairwise}\ \ MP)$$

  In Gaussian case
  $k_{uv} = 0 \iff X_u \perp\!\!\!\perp X_v \mid \mathbf{X}_{-uv} = (X_w,\ w \neq u \text{ and } v)'$.
- A, B and S three disjoints subsets of V :

$$S \text{ separates } A \text{ and } B \ \Rightarrow\ \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S \quad (\underline{Global}\ \ MP)$$

Graphical Models : Perfect Markov distributions

- In Gaussian case : (Pairwise) $\Longleftrightarrow$ (Global)
- The inverse in (Global) is not always true :

Graphical Models : Perfect Markov distributions

- In Gaussian case : (Pairwise) $\iff$ (Global)
- The inverse in (Global) is not always true :

$$\Sigma = \begin{pmatrix} 4 & -1 & -1 & -1 \\ -1 & 2 & 0 & 0 \\ -1 & 0 & 2 & 0 \\ -1 & 0 & 0 & 2 \end{pmatrix}, \; K = \begin{pmatrix} 0.4 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.6 & 0.1 & 0.1 \\ 0.2 & 0.1 & 0.6 & 0.1 \\ 0.2 & 0.1 & 0.1 & 0.6 \end{pmatrix}$$

$$\Sigma_{\{2,3,4\}} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \; (\Sigma_{\{2,3,4\}})^{-1} = \begin{pmatrix} .5 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .5 \end{pmatrix}$$

$$G = G(K) \text{ is complete}, \quad X_2 \perp\!\!\!\perp X_3 \mid X_4$$

Definition

The distribution P of a r.v. $\mathbf{X}_V$ is <u>*Perfectly*</u> Markov to G if

$$S \text{ separates } A \text{ and } B \iff \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S$$

Definition

The distribution P of a r.v. $\mathbf{X}_V$ is *Perfectly* Markov to G if

$$S \text{ separates } A \text{ and } B \iff \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S$$

Theorem, Meek 1995

For any undirected graph $G = (V, E)$, $\exists\ P = \mathcal{N}_{|V|}(\mu, K = \Sigma^{-1})$ such that $G(K) = G$ and P is perfectly Markov to G.

Theorem, Geiger *et al.* 2000

If $G = G(K)$ is a tree, then for all $P = \mathcal{N}_{|V|}(\mu, K = \Sigma^{-1})$ such that $G(K) = G$, P is perfectly Markov to G.

Problem :

- $X^{(1)}, \dots, X^{(N)}$ i.i.d. $P = \mathcal{N}_{|V|}(\mu, K = \Sigma^{-1})$.
  Estimate $G = G(K)$ in the case $N \leq |V|$, No MLE of K... ?

- $X^{(1)}, \dots, X^{(N)}$ i.i.d. $P = \mathcal{N}_{|V|}(\mu, K = \Sigma^{-1})$.
  Estimate $G = G(K)$ in the case $N \leq |V|$, No MLE of K... ?

- *Proposed solution* : Naive PC-algorithm

  1. test marginal independence : $H_0^{uv} \;:\; X_u \perp\!\!\!\perp X_v, \; \forall\, u, \, v.$
  2. test conditional independence given 1 variable

     $$H_0^{uv|w} \;:\; X_u \perp\!\!\!\perp X_v \mid X_w, \; \forall\, u, \, v, \, w$$

  3. test conditional independence given 2 variables

     $$H_0^{uv|S} \;:\; X_u \perp\!\!\!\perp X_v \mid \mathbf{X}_S, \; \forall\, u, \, v, \text{ and } S \subseteq V \setminus \{u, v\}, \; |S| = 2$$

  .....and so on.

Questions ??

- Do I obtain the true graph ?
- What is the needed number of conditioning variables ?
- Should I test on all the subsets with cardinality k ?

PC-algorithms and Variations

- $0-1-$procedure : Marginal Independence and Condition on 1 variable (Friedman *et al* 2000, Magwene and Kim 2004, Wille and Bühlman 2006)
- PC-algorithm and variations :
  - Skeleton Bayesian Networks : Verma and Pearl 1991, Steck and Tresp 1999, Spirtes *et al*. 2000, Kalish and Bühlman 2007...
  - Undirected graphs : Castello and Roverato 2006 (qp-procedure), Malouche and Sevestre 2008 (uPC-algorithm).

$k-$graphs

Definition

$\mathbf{X} = (X_u,\ u \in V)' \sim\ P = \mathcal{N}_{|V|}(\mu, K = \Sigma^{-1})$

$G_k = G_k(P) = (V, E_k)$ is called a $\underline{k-graph}$, if

$$u \not\sim_{G_k} v \quad \Longleftrightarrow \quad \begin{cases} \exists\, S \subseteq V \setminus \{u, v\},\ |S| = k, \\ X_u \perp\!\!\!\perp X_v \mid \mathbf{X}_S \end{cases}$$

$k-$graphs

Definition

$\mathbf{X} = (X_u,\ u \in V)' \sim\ P = \mathcal{N}_{|V|}(\mu, K = \Sigma^{-1})$

$G_k = G_k(P) = (V, E_k)$ is called a *k−graph*, if

$$u \not\sim_{G_k} v \quad \Longleftrightarrow \quad \begin{cases} \exists\, S \subseteq V \setminus \{u,\ v\},\ |S| = k, \\ X_u \perp\!\!\!\perp X_v \mid \mathbf{X}_S \end{cases}$$

Examples.

- $0-$graph : $G_0(P) = G(\Sigma)$ is the *covariance graph* (or bi-directed graph)
- $(|V| - 2)-$graph : $G_{|V|-2}(P) = G(K)$ is the *concentration graph*

Separability order

$G = (V, E)$ an undirected graph

Definition

Separability order of G : if $u \not\sim_G v$

$$\mathrm{so}(u, v, G) = \min \{ |S|, \; S \text{ is a separator of } u \text{ and } v \}$$

$$\mathrm{so}(G) = \max_{u \not\sim_G v} \mathrm{so}(u, v, G)$$

Remarks.

- $\mathrm{so}(u, v, G) = 0 \iff$ u and v belong to different connected components
- $\mathrm{so}(G) = 0 \iff G$ is union of disconnected complete graphs.

Theorem, M¬ and Sevestre 2008

$\mathbf{X} = (X_u,\ u \in V)' \sim P = \mathcal{N}_{|V|}(\mu, K = \Sigma^{-1})$.

$G = G(K)$ the concentration graph

$G_k = G(P)$ the $k-$graph

Assume that

i. P is perfectly Markov to G

ii. $\mathrm{so}(G) = m$

iii. $\mathrm{so}(G_0) < |V| - 2$.

Then

$$G = G_m \subseteq G_{m-1} \subseteq \ldots G_1 \subseteq G_0$$

Remarks

- uPC-theorem remains true for categorical multivariate r.v.
- $G_1 \subseteq G_0$ can be obtained using Global Markov Property on G_0.
- uPC-procedure : Estimate $\widehat{G}_0$, then $\widehat{G}_1, \ldots, \widehat{G}_k$ till $\mathrm{so}(\widehat{G}_k) = k$.
- *Questions* :
  - Can we check the no Perfect Markovianity from G_0 ?
  - Which conditioning subset of variables should I consider ?

Covariance Decomposition

Theorem, Jones and West 2005

$\mathbf{X} = (X_u,\ u \in V)' \sim P = \mathcal{N}_{|V|}(\mu, K = \Sigma^{-1})$.

$G = G(K)$ the concentration graph

$G_0 = G_0(P) = G(\Sigma)$ the $0-$graph

For all u and v in V

$$k_{uv} = \sum_{p=p(u,v,G_0)} (-1)^{|p|+1} \sigma_p \frac{|\Sigma \setminus p|}{|\Sigma|}$$

and

$$\sigma_{uv} = \sum_{p=p(u,v,G)} (-1)^{|p|+1} k_p \frac{|K \setminus p|}{|K|}$$

where $\sigma_p = \sigma_{u_0 u_1} \dots \sigma_{u_{n-1} u_n}$ and $k_p = k_{u_0 u_1} \dots k_{u_{n-1} u_n}$ if $p = (u_0, \dots, u_n)$.

Consequences

- If G is a tree then G_0 is complete
- If G_0 is a tree then G is complete, and P could not be Perfectly Markov to G.

Consequences

- If G is a tree then G_0 is complete
- If G_0 is a tree then G is complete, and P could not be Perfectly Markov to G.

Proposition

$\mathbf{X} = (X_u,\ u \in V)' \sim P = \mathcal{N}_{|V|}(\mu, K = \Sigma^{-1})$.

$G = G(K)$ the concentration graph

$G_0 = G_0(P) = G(\Sigma)$ the $0-$graph

If $\exists\ u \not\sim_{G_0} v$ and $\exists!\ p(u, v, G_0)$, $|p| \geq 2$
then P is not perfectly Markov to G.

Proof

Proof.

1. If P is perfectly Markov to G, then $G \subseteq G_0$
2. $u \not\sim_{G_0} v$, then $u \not\sim_G v$ and $k_{uv} = 0$
3. If $p = p(u, v, G_0) = (u_0, \ldots, u_n)$ is the path connecting u and v in G_0, then

$$k_{uv} = (-1)^{n+1} \sigma_{uu_1} \ldots \sigma_{u_{n-1}v} \frac{|\Sigma \setminus p|}{|\Sigma|} \neq 0$$

Contradiction.

□

Subset of conditioning variables, 1st Result

For all $u, v \in V$

$$T(u, v, G) = \{w \in V \setminus \{u, v\},\ \exists\, p(u, w, G) \not\ni v \text{ and } p(w, v, G) \not\ni u\}$$

Subset of conditioning variables, 1st Result

For all $u, v \in V$

$$T(u, v, G) = \{w \in V \setminus \{u, v\},\ \exists\, p(u, w, G) \not\ni v \text{ and } p(w, v, G) \not\ni u\}$$

Theorem

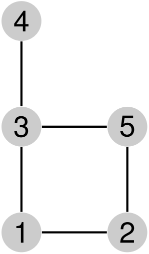
$\mathbf{X} = (X_u,\ u \in V)' \sim P = \mathcal{N}_{|V|}(\mu, K = \Sigma^{-1})$.

- G_0 is the $0-$graph.
- $u, v \in V$, $S \subseteq V \setminus \{u, v\}$

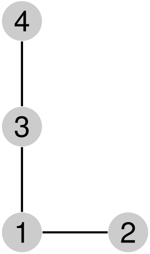
Assume that $S \cap T\left(u, v, (G_0)_{(S \cup \{u,v\})}\right) = \emptyset$, then

$$X_u \perp\!\!\!\perp X_v \iff X_u \perp\!\!\!\perp X_v \mid \mathbf{X}_S$$

$$\mathbf{X} = (X_1, X_2, X_3, X_4, X_5)'$$



G_0



$(G_0)_{\{1,2,3,4\}}$

$S = \{3, 4\}$ and $u = 1$, $v = 2$

$$k_{12|34} = 0 \iff \sigma_{12} = 0$$

Proof

Proof.

1. If $X_u \perp\!\!\!\perp X_v$, Condition $S \cap T\left(u, v, (G_0)_{(S\cup\{u,v\})}\right) = \emptyset$ implies

   $(G_0)_{(S\cup\{u,v\})}$ contains at least two connected components $S_1 \ni u$ and $S_2 \ni v$.

   $\Rightarrow$ $k_{uv|S} = 0$.

2. If $X_u \not\perp\!\!\!\perp X_v$, then $\sigma_{uv} \neq 0$, and

$$k_{uv|S} = (-1)^{1+1}\sigma_{uv}\frac{|\Sigma_S|}{|\Sigma_{(S\cup\{u,v\})}|} \neq 0.$$

$\square$

Theorem

$\mathbf{X} = (X_u,\ u \in V)' \sim P = \mathcal{N}_{|V|}(\mu, K = \Sigma^{-1})$.
G_0 is the 0−graph, $u, v \in V$, $S \subseteq V \setminus \{u, v\}$, and $w \in V \setminus (S \cup \{u, v\})$.
If

i. $w \notin \mathrm{ne}_{G_0}(u) \cap \mathrm{ne}_{G_0}(v)$
ii. $w \notin \mathrm{ne}_{G_0}(S)$

Then

$$X_u \perp\!\!\!\perp X_v \mid \mathbf{X}_S \iff X_u \perp\!\!\!\perp X_v \mid \mathbf{X}_{S \cup \{w\}}$$

Proof

Proof.

W.l.g. assume $w \sim_{G_0} u$ and $w \not\sim_{G_0} v$.

$$\Sigma_{S\cup\{u,w\}\times S\cup\{v,w\}} = \begin{pmatrix} \sigma_{ww} & 0 & 0 & \dots & 0 \\ \sigma_{wu} & & & & \\ 0 & & \Sigma_{S\cup\{u\}\times S\cup\{v\}} & & \\ 0 & & & & \end{pmatrix}$$

Then

$$\left|\Sigma_{S\cup\{u,w\}\times S\cup\{u,w\}}\right| = \sigma_{ww}\left|\Sigma_{S\cup\{u\}\times S\cup\{v\}}\right|$$

□

Subset of conditioning variables

Corollary

$\mathbf{X} = (X_u,\ u \in V)' \sim P = \mathcal{N}_{|V|}(\mu, K = \Sigma^{-1})$ Perfectly Markov to $G = G(K)$.

- $u, v \in V$, $u \sim_{G_{k-1}} v$
- $S \subseteq V \setminus \{u, v\}$, $|S| = k$.

Assume that $\exists\, w \in S$ satisfying

i. $w \notin \mathrm{ne}_{G_0}(u) \cap \mathrm{ne}_{G_0}(v)$

ii. $w \notin \mathrm{ne}_{G_0}(S \setminus \{w\})$

Then

$$X_u \not\perp\!\!\!\perp X_v \mid \mathbf{X}_S$$

A last result !

Definition

Let $G \subseteq G_0$. $P_G \subseteq V$ defined as follows :

$$u \in P_G \quad \Longleftrightarrow \quad \exists \text{ a path } p \ni u, |p| \geq 2 \text{ and } \left\{ \begin{array}{l} G \setminus p \text{ disconnected} \\ p \ni w,\ w \not\sim_{G_0} u \end{array} \right.$$

Proposition

$\mathbf{X} = (X_u,\ u \in V)' \sim P = \mathcal{N}_{|V|}(\mu, K = \Sigma^{-1})$.
P is Perfectly Markov to $G = G(K)$ and G_0 connected. Let $u, v \in V$, $u \sim_{G_{k-1}} v$.
Assume that u or $v \in P_{G_{k-1} \setminus \{u,v\}}$. Then

$$u \sim_{G_k} v$$

Proof

Proof.

Assume $u \not\sim_{G_k} v$ and $u \in P_{G_{k-1} \setminus \{u,v\}}$:
$p(u, w, G_{k-1} \setminus (u, v)) = (u_0, \ldots, u_n)$

- If $\exists\, i$, such that $u_i \not\sim_G u_{i+1}$. Then G becomes disconnected (G connected $\iff$ G_0 connected) : Contradiction.
- Then $p(u, w, G_{k-1} \setminus (u, v)) = p(u, w, G) \Rightarrow \sigma_{uw} \neq 0$.

Contradiction. $\square$

So which are the remaining conditioning subsets ?

$u \sim_{G_{k-1}} v$

$$\mathcal{S}_k(u,v) = \{S \subseteq V \setminus \{u,v\},\ |S| = k,\ \text{such that } X_u \perp\!\!\!\perp X_v \mid \mathbf{X}_S \text{ could not be deduced from the step } k-1\}$$

So which are the remaining conditioning subsets ?

$u \sim_{G_{k-1}} v$

$$\mathcal{S}_k(u,v) = \{S \subseteq V \setminus \{u,v\},\ |S| = k,\ \text{such that } X_u \perp\!\!\!\perp X_v \mid \mathbf{X}_S \text{ could not be deduced from the step } k-1\}$$

Proposition

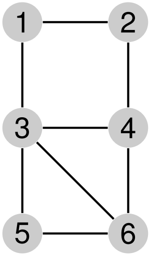
Let $u \sim_{G_{k-1}} v$. Assume that u and v are $\notin P_{G_{k-1} \setminus \{u,v\}}$. Let $S \subseteq V \setminus \{u,v\}$ and $|S| = k$.

$$S \in \mathcal{S}_k(u,v) \iff S = T(u,v,(G_0)_{S \cup \{u,v\}})$$

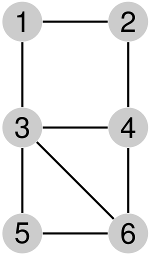
Proof.

By induction on k. □

Example

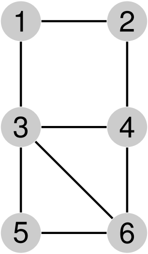


Example



$\mathrm{so}(G_0) = 2$

Example



$\mathrm{so}(G_0) = 2$

CI Tests

$3 \perp\!\!\!\perp 4 \mid \{6\}$

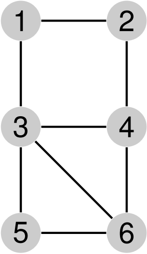
$3 \perp\!\!\!\perp 5 \mid \{6\}$

$3 \perp\!\!\!\perp 6 \mid \{4\}, 3 \perp\!\!\!\perp 6 \mid \{5\}$

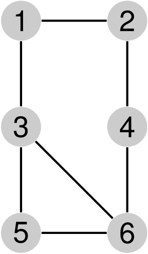
$4 \perp\!\!\!\perp 6 \mid \{3\}$

$5 \perp\!\!\!\perp 6 \mid \{3\}$

Example



$\mathrm{so}(G_0) = 2$



CI Tests

$3 \perp\!\!\!\perp 4 \mid \{6\}$

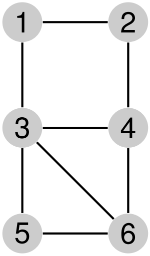
$3 \perp\!\!\!\perp 5 \mid \{6\}$

$3 \perp\!\!\!\perp 6 \mid \{4\}, 3 \perp\!\!\!\perp 6 \mid \{5\}$

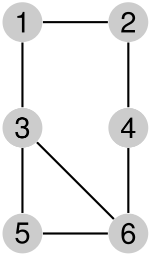
$4 \perp\!\!\!\perp 6 \mid \{3\}$

$5 \perp\!\!\!\perp 6 \mid \{3\}$

Example



$\mathrm{so}(G_0) = 2$



$\mathrm{so}(G_1) = 2$

CI Tests

$3 \perp\!\!\!\perp 4 \mid \{6\}$

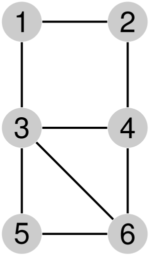
$3 \perp\!\!\!\perp 5 \mid \{6\}$

$3 \perp\!\!\!\perp 6 \mid \{4\}, 3 \perp\!\!\!\perp 6 \mid \{5\}$

$4 \perp\!\!\!\perp 6 \mid \{3\}$

$5 \perp\!\!\!\perp 6 \mid \{3\}$

Example



$\mathrm{so}(G_0) = 2$

CI Tests

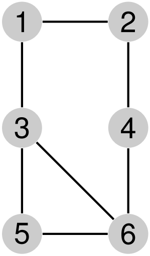
$3 \perp\!\!\!\perp 4 \mid \{6\}$

$3 \perp\!\!\!\perp 5 \mid \{6\}$

$3 \perp\!\!\!\perp 6 \mid \{4\}$, $3 \perp\!\!\!\perp 6 \mid \{5\}$

$4 \perp\!\!\!\perp 6 \mid \{3\}$

$5 \perp\!\!\!\perp 6 \mid \{3\}$



$\mathrm{so}(G_1) = 2$

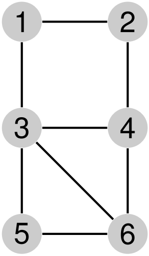
CI Tests

$3 \perp\!\!\!\perp 5 \mid \{4, 6\}$

$3 \perp\!\!\!\perp 6 \mid \{4, 5\}$, $3 \perp\!\!\!\perp 6 \mid \{4, 2\}$

$5 \perp\!\!\!\perp 6 \mid \{3, 4\}$, $5 \perp\!\!\!\perp 6 \mid \{3, 1\}$

Example



$\mathrm{so}(G_0) = 2$

CI Tests

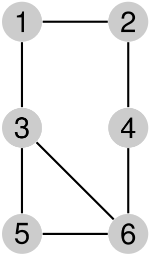
$3 \perp\!\!\!\perp 4 \mid \{6\}$

$3 \perp\!\!\!\perp 5 \mid \{6\}$

$3 \perp\!\!\!\perp 6 \mid \{4\}, 3 \perp\!\!\!\perp 6 \mid \{5\}$

$4 \perp\!\!\!\perp 6 \mid \{3\}$

$5 \perp\!\!\!\perp 6 \mid \{3\}$



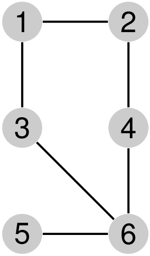
$\mathrm{so}(G_1) = 2$

CI Tests

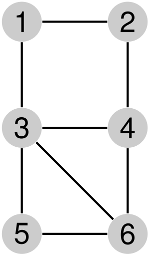
$3 \perp\!\!\!\perp 5 \mid \{4, 6\}$

$3 \perp\!\!\!\perp 6 \mid \{4, 5\}, 3 \perp\!\!\!\perp 6 \mid \{4, 2\}$

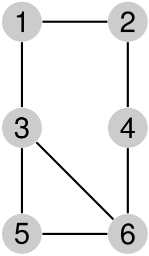
$5 \perp\!\!\!\perp 6 \mid \{3, 4\}, 5 \perp\!\!\!\perp 6 \mid \{3, 1\}$



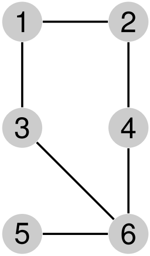
Example



$\mathrm{so}(G_0) = 2$



$\mathrm{so}(G_1) = 2$



$\mathrm{so}(G_2) = 2$

CI Tests

$3 \perp\!\!\!\perp 4 \mid \{6\}$

$3 \perp\!\!\!\perp 5 \mid \{6\}$

$3 \perp\!\!\!\perp 6 \mid \{4\},\ 3 \perp\!\!\!\perp 6 \mid \{5\}$

$4 \perp\!\!\!\perp 6 \mid \{3\}$

$5 \perp\!\!\!\perp 6 \mid \{3\}$

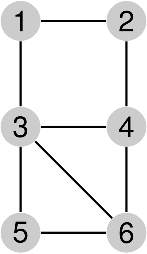
CI Tests

$3 \perp\!\!\!\perp 5 \mid \{4, 6\}$

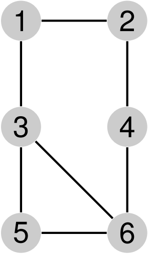
$3 \perp\!\!\!\perp 6 \mid \{4, 5\},\ 3 \perp\!\!\!\perp 6 \mid \{4, 2\}$

$5 \perp\!\!\!\perp 6 \mid \{3, 4\},\ 5 \perp\!\!\!\perp 6 \mid \{3, 1\}$

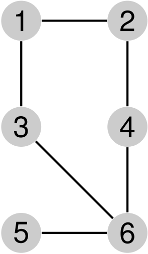
Example



$\mathrm{so}(G_0) = 2$



$\mathrm{so}(G_1) = 2$



$\mathrm{so}(G_2) = 2$

CI Tests

$3 \perp\!\!\!\perp 4 \mid \{6\}$

$3 \perp\!\!\!\perp 5 \mid \{6\}$

$3 \perp\!\!\!\perp 6 \mid \{4\}, 3 \perp\!\!\!\perp 6 \mid \{5\}$

$4 \perp\!\!\!\perp 6 \mid \{3\}$

$5 \perp\!\!\!\perp 6 \mid \{3\}$

CI Tests

$3 \perp\!\!\!\perp 5 \mid \{4, 6\}$

$3 \perp\!\!\!\perp 6 \mid \{4, 5\}, 3 \perp\!\!\!\perp 6 \mid \{4, 2\}$

$5 \perp\!\!\!\perp 6 \mid \{3, 4\}, 5 \perp\!\!\!\perp 6 \mid \{3, 1\}$

$\#$ of Tests =26