Optimal designs for state estimation in time-dependent networks

Justin Lamberti, Prof. Dr. Kirsten Schorning, Prof. Dr. Christine Müller

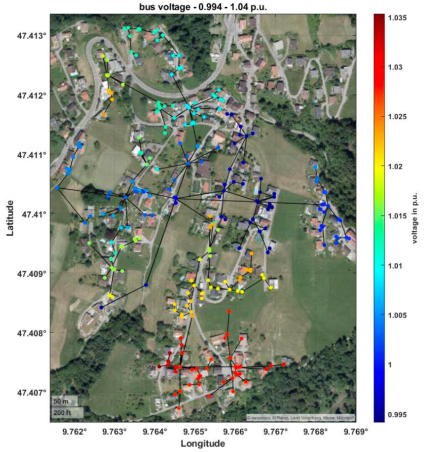
mODa 14

June 18th, 2026, Dongen Abbey, Belgium

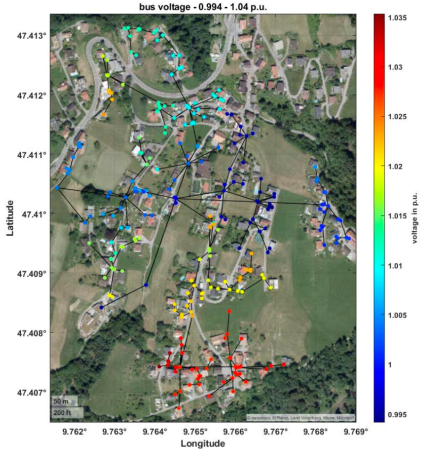
technische universität dortmund

Transregio 391

Motivating Example



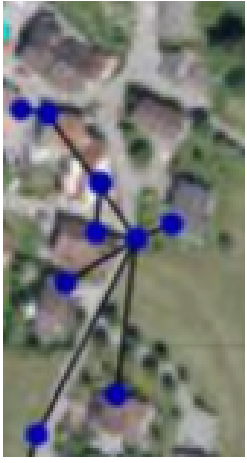
Motivating Example



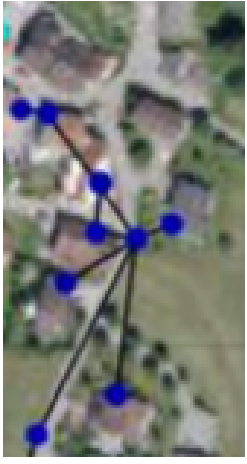
Goals:

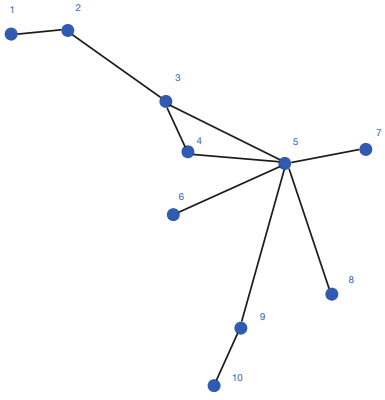
- Build an appropriate statistical model
- Study different designs and their performance on networks

Motivation

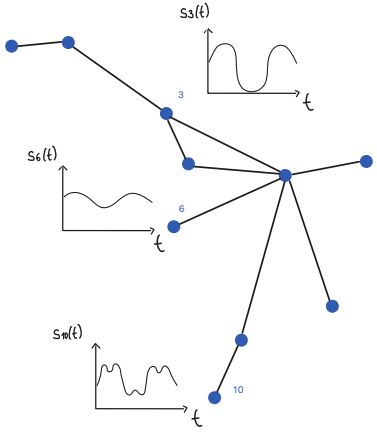


Motivation

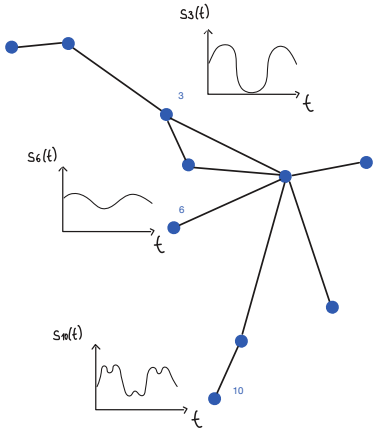




Different periodical behavior in each node



Different periodical behavior in each node



Goal:

Estimate $s_1(t), \ldots, s_{10}(t)$ using observations at the different nodes that are collected by different measurement devices.

Challenges while building the model

- **First challenge:** States in the nodes are not deterministic, but random, thus we obtain

$$S_j(t) = s_j(t) + Z_j(t) \quad \text{for } j \in \{1, \ldots, I\}.$$

Challenges while building the model

- **First challenge:** States in the nodes are not deterministic, but random, thus we obtain

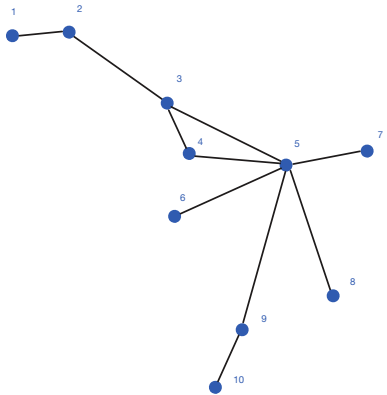
$$S_j(t) = s_j(t) + Z_j(t) \quad \text{for } j \in \{1, \ldots, I\}.$$

- **Second challenge:** Observations at the j-th node are not only influenced by the state $s_j(t)$, but also the of the adjacent nodes, thus

$$Y_{jn} = S_j(t) + \sum_{k \in A_j} \alpha_k S_k(t) + \mathbf{e}_{jn}$$

with A_j being the set of adjacent nodes to the j-th node.

Research Question



Question of interest:

How to allocate the different measurement devices at the nodes of the network to obtain precise estimations for $s_1(t), \ldots, s_l(t)$?

Building the model

The model

We consider the random effects model with time-dependent states as

$$\mathbf{Y}_n = \mathbb{X}\mathbf{s}(t_n) + \mathbb{X}\mathbf{Z}_n + \mathbf{e}_n \in \mathbb{R}^I$$

- Heteroscedastic errors $\mathrm{Cov}(\mathbf{Y}_n) = \sigma^2(\mathbb{X}\mathbf{D}_Z\mathbb{X}^\top + \mathbf{D}_e) := \sigma^2\mathbf{S}$ with $\mathbf{D}_e = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_I^2)$ and $\mathbf{D}_Z$ positive semidefinite matrix.
- State vector $\mathbf{s}(t_n) \in \mathbb{R}^I$ at time point t_n.
- Influence matrix $\mathbb{X} \in \mathbb{R}^{I\times I}$.

The model

We consider the random effects model with time-dependent states as

$$\mathbf{Y}_n = \mathbb{X}\mathbf{s}(t_n) + \mathbb{X}\mathbf{Z}_n + \mathbf{e}_n$$

- Heteroscedastic errors $\mathrm{Cov}(\mathbf{Y}_n) = \sigma^2(\mathbb{X}\mathbf{D}_Z\mathbb{X}^\top + \mathbf{D}_e) := \sigma^2\mathbf{S}$ with $\mathbf{D}_e = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_I^2)$ and $\mathbf{D}_Z$ positive semidefinite matrix.
- State vector $\mathbf{s}(t_n) \in \mathbb{R}^I$ at time point t_n.
- Influence matrix $\mathbb{X} \in \mathbb{R}^{I \times I}$.
- **Assumption 1**: $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$ can be treated as mutually independent.
- **Assumption 2**: $s_j(t) = \mathbf{f}^\top(t)\boldsymbol{\theta}_j \ \forall j \in \{1, \ldots, I\}$

Idea for the second assumption given no prior knowledge

Assumption 2: $s_j(t) = \mathbf{f}^\top(t)\boldsymbol{\theta}_j \ \forall j \in \{1, \ldots, I\}$.

The idea under no prior knowledge:

Every complex periodical function can be rewritten as a Fourier-Series of the following form:

$$s_j(t) \sim \theta_{0j} + \sum_{k=1}^{\infty} \left[\theta_{kj}^{(1)} \cos\left(\frac{2\pi kt}{T}\right) + \vartheta_{kj}^{(2)} i \sin\left(\frac{2\pi kt}{T}\right) \right],$$

for $t \in [0, T]$.

Idea for the second assumption given no prior knowledge

Assumption 2: $s_j(t) = \mathbf{f}^\top(t)\boldsymbol{\theta}_j \ \forall j \in \{1, \ldots, I\}$.

The idea under no prior knowledge:

Every real-valued periodical function can be rewritten as a Fourier-Series of the following form:

$$s_j(t) \sim \theta_{0j} + \sum_{k=1}^{\infty} \left[\theta_{kj}^{(1)} \cos\left(\frac{2\pi kt}{T}\right) + \theta_{kj}^{(2)} \sin\left(\frac{2\pi kt}{T}\right) \right],$$

for $t \in [0, T]$.

Idea for the second assumption given no prior knowledge

Assumption 2: $s_j(t) = \mathbf{f}^\top(t)\boldsymbol{\theta}_j \ \forall j \in \{1, \dots, I\}$.

The idea under no prior knowledge:

Every real-valued periodical function can be rewritten as a Fourier-Series of the following form:

$$s_j(t) \sim \theta_{0j} + \sum_{k=1}^{\infty} \left[\theta_{kj}^{(1)} \cos\left(\frac{2\pi kt}{T}\right) + \theta_{kj}^{(2)} \sin\left(\frac{2\pi kt}{T}\right) \right],$$

for $t \in [0, T]$. We cut off the Fourier Series after m summands, and obtain:

- $\boldsymbol{\theta}_j = (\theta_{0j}, \theta_{1j}^{(1)}, \dots, \theta_{mj}^{(1)}, \theta_{1j}^{(2)}, \dots, \theta_{mj}^{(2)})^\top \in \mathbb{R}^{2m+1}$
- $\mathbf{f}^\top(t) = (1, \cos(\frac{2\pi t}{T}), \dots, \cos(\frac{2\pi mt}{T}), \sin(\frac{2\pi t}{T}), \dots, \sin(\frac{2\pi mt}{T}))$

Model building for multiple points in time

Under linear functional relationship we obtain:

$$\mathbf{Y}_n = \left(\mathbb{X} \otimes \mathbf{f}^\top(t_n)\right)\boldsymbol{\theta} + \mathbb{X}\mathbf{Z}_n + \mathbf{e}_n$$

If $p \leq N$ holds, this results in:

$$\hat{\mathbf{s}}(t) = \sum_{j=1}^{N} \left(\mathbf{f}^\top(t) \left(\sum_{k=1}^{N} \mathbf{f}(t_k)\mathbf{f}^\top(t_k) \right)^{-} \mathbf{f}(t_j) \right) \left((\mathbb{X}^\top \mathbf{S}^{-1}\mathbb{X})^{-}\mathbb{X}^\top \mathbf{S}^{-1}\mathbf{Y}_j \right).$$

The covariance structure is

$$\mathrm{Cov}(\hat{\mathbf{s}}(t)) = \sigma^2 (\mathbf{f}^\top(t) \left(\sum_{k=1}^{N} \mathbf{f}(t_k)\mathbf{f}^\top(t_k) \right)^{-} \mathbf{f}(t)) \cdot (\mathbb{X}^\top \mathbf{S}^{-1}\mathbb{X})^{-}.$$

Separability of the covariance matrix

The covariance of the estimator is given by

$$\mathrm{Cov}(\hat{\mathbf{s}}(t)) = \sigma^2(\mathbf{f}^\top(t)\left(\sum_{k=1}^{N}\mathbf{f}(t_k)\mathbf{f}^\top(t_k)\right)^{-}\mathbf{f}(t))\left(\mathbb{X}^\top\mathbf{S}^{-1}\mathbb{X}\right)^{-}.$$

Separability of the covariance matrix

The covariance of the estimator is given by

$$\mathrm{Cov}(\hat{\mathbf{s}}(t)) = \sigma^2(\mathbf{f}^\top(t)\left(\sum_{k=1}^{N}\mathbf{f}(t_k)\mathbf{f}^\top(t_k)\right)^{-}\mathbf{f}(t))\left(\mathbb{X}^\top\mathbf{S}^{-1}\mathbb{X}\right)^{-}.$$

- $\mathbf{f}^\top(t)\left(\sum_{k=1}^{N}\mathbf{f}(t_k)\mathbf{f}^\top(t_k)\right)^{-}\mathbf{f}(t) \in \mathbb{R}$ represents the **temporal** part.

Separability of the covariance matrix

The covariance of the estimator is given by

$$\mathrm{Cov}(\hat{\mathbf{s}}(t)) = \sigma^2(\mathbf{f}^\top(t)\left(\sum_{k=1}^{N}\mathbf{f}(t_k)\mathbf{f}^\top(t_k)\right)^{-}\mathbf{f}(t))\left(\mathbb{X}^\top\mathbf{S}^{-1}\mathbb{X}\right)^{-}.$$

- $\mathbf{f}^\top(t)\left(\sum_{k=1}^{N}\mathbf{f}(t_k)\mathbf{f}^\top(t_k)\right)^{-}\mathbf{f}(t) \in \mathbb{R}$ represents the **temporal** part.
- $\mathbb{X}^\top\mathbf{S}^{-1}\mathbb{X} \in \mathbb{R}^{I\times I}$ represents the **spatial** part.

Separability of the covariance matrix

The covariance of the estimator is given by

$$\mathrm{Cov}(\hat{\mathbf{s}}(t)) = \sigma^2 (\mathbf{f}^\top(t) \left(\sum_{k=1}^{N} \mathbf{f}(t_k)\mathbf{f}^\top(t_k) \right)^{-} \mathbf{f}(t)) \left(\mathbb{X}^\top \mathbf{S}^{-1} \mathbb{X} \right)^{-}.$$

- $\mathbf{f}^\top(t)\left(\sum_{k=1}^{N} \mathbf{f}(t_k)\mathbf{f}^\top(t_k)\right)^{-}\mathbf{f}(t) \in \mathbb{R}$ represents the **temporal** part.
- $\mathbb{X}^\top \mathbf{S}^{-1} \mathbb{X} \in \mathbb{R}^{I \times I}$ represents the **spatial** part.

$\Rightarrow$ We can optimize the **spatial** and **temporal** parts independently of each other.

DoE on networks

Derive a Design of Experiment Problem

If we assume that $\mathbb{X} \in \mathbb{R}^{I \times I}$ is invertible, the spatial part of the covariance is given by

$$(\mathbb{X}^\top \mathbf{S}^{-1} \mathbb{X})^- = (\mathbb{X}^\top (\mathbb{X} \mathbf{D}_Z \mathbb{X}^\top + \mathbf{D}_e)^{-1} \mathbb{X})^- = \mathbf{D}_Z + \mathbb{X}^{-1} \mathbf{D}_e (\mathbb{X}^{-1})^\top,$$

with $\mathbf{D}_e = \text{diag}(\sigma_1^2, \ldots, \sigma_I^2)$ containing variances according to different measuring procedures.

Derive a Design of Experiment Problem

If we assume that $\mathbb{X} \in \mathbb{R}^{I \times I}$ is invertible, the spatial part of the covariance is given by

$$(\mathbb{X}^\top \mathbf{S}^{-1} \mathbb{X})^- = (\mathbb{X}^\top (\mathbb{X} \mathbf{D}_Z \mathbb{X}^\top + \mathbf{D}_e)^{-1} \mathbb{X})^- = \mathbf{D}_Z + \mathbb{X}^{-1} \mathbf{D}_e (\mathbb{X}^{-1})^\top,$$

with $\mathbf{D}_e = \text{diag}(\sigma_1^2, \ldots, \sigma_I^2)$ containing variances according to different measuring procedures.

$\Rightarrow$ The precision of the measuring procedure in the j-th node is given by $\delta_j = \frac{1}{\sigma_j^2}$, in matrix notation $\mathbf{D}_e^{-1} = \mathbf{D}_\delta$.

Derive a Design of Experiment Problem

If we assume that $\mathbb{X} \in \mathbb{R}^{I \times I}$ is invertible, the spatial part of the covariance is given by

$$(\mathbb{X}^\top \mathbf{S}^{-1} \mathbb{X})^- = (\mathbb{X}^\top (\mathbb{X} \mathbf{D}_Z \mathbb{X}^\top + \mathbf{D}_e)^{-1} \mathbb{X})^- = \mathbf{D}_Z + \mathbb{X}^{-1} \mathbf{D}_e (\mathbb{X}^{-1})^\top,$$

with $\mathbf{D}_e = \text{diag}(\sigma_1^2, \ldots, \sigma_I^2)$ containing variances according to different measuring procedures.

$\Rightarrow$ The precision of the measuring procedure in the j-th node is given by $\delta_j = \frac{1}{\sigma_j^2}$, in matrix notation $\mathbf{D}_e^{-1} = \mathbf{D}_\delta$.

$\Rightarrow$ We obtain design space for the approximate designs as

$$\delta \in \Delta := \left\{ \delta = (\delta_1, \ldots, \delta_I)^\top \in (0,1)^I ; \sum_{j=1}^{I} \delta_j = 1 \right\}.$$

If $\mathbb{X}$ is invertible, the optimization problem is given by:

$$\delta^* = \arg\min_{\delta \in \Delta} \phi_A(\delta) = \arg\min_{\delta \in \Delta} \left\{ \text{tr}((\mathbb{X}^\top \mathbf{D}_\delta \mathbb{X})^{-1}) \right\}.$$

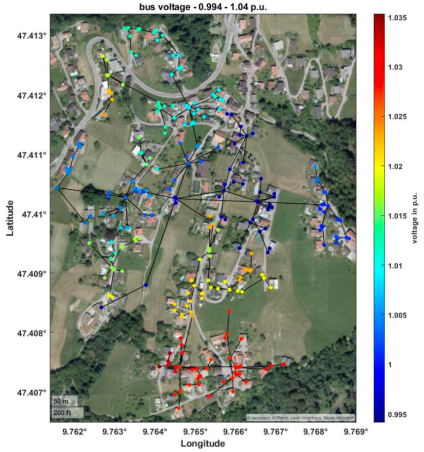
An analytical solution

$\delta^* = (\delta_1^*, \ldots, \delta_I^*)$ is an A-optimal design and therefore a solution to the DoE problem if and only if

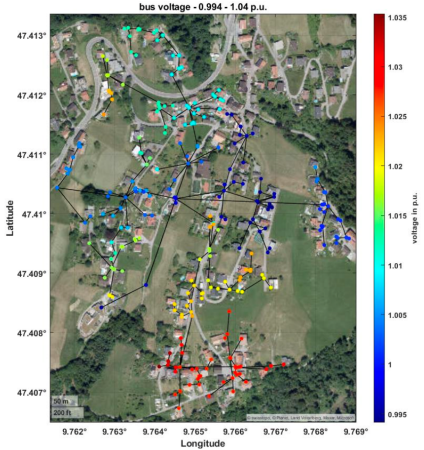
$$\delta_j^* = \frac{\sqrt{v_j}}{\sum_{k=1}^{I} \sqrt{v_k}}, \text{ with } v_j = \mathbf{u}_j^\top (\mathbb{X}^{-1})^\top \mathbb{X}^{-1} \mathbf{u}_j \text{ for } j \in \{1, \ldots, I\},$$

where $\mathbf{u}_j$ denotes the j-th unit vector.

Is the application feasible for distribution grids?



Is the application feasible for distribution grids?



Goals:

- Inverting large matrices computationally expensive
- Inverting large matrices computationally unstable

The designs

The following designs are considered:

1. The optimal design δ^*.
2. The equidistant design $\delta_{eq} = (\frac{1}{l}, \ldots, \frac{1}{l})$.
3. The degree design $\delta_{deg} \propto (deg(1), \ldots, deg(l))$.
4. The design based on the quotient graph method δ_{quot}.
5. The design based on the developed deterministic cutting algorithm δ_{cut}.

Reducing networks

Example for the skeleton and the quotient graph

The automorphism group is given by $Aut(\mathcal{G}) := \{\eta \mid \eta(\mathcal{E}) = \mathcal{E}\}$.

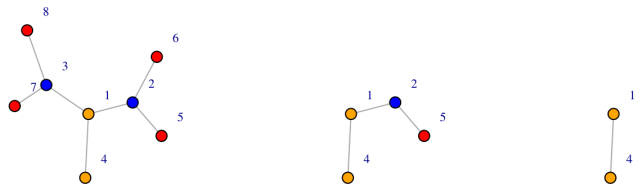


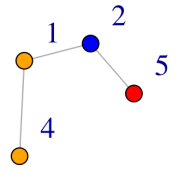
Figure 1: On the left-hand side a graph, in the middle the quotient graph and on the right-hand side the skeleton of the corresponding graph.

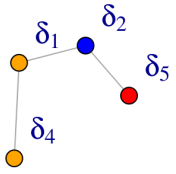
The unique vertex orbits are given as $\{\{1\}, \{4\}, \{2, 3\}, \{5, 6, 7, 8\}\}$.

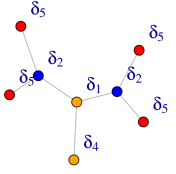
Design based on the quotient graph

The procedure for obtaining a design based on the quotient graph:

- Calculate the quotient graph of the supergraph.
- Calculate the optimal design on the optimal graph.
- Give the vertices the same weight as the vertices in their respective vertex orbit.
- Rescale the design.







Example of the deterministic cutting algorithm

The cutting algorithm cuts the supergraph into smaller subgraphs. The supergraph is to cut along the skeletons to build subgraphs, which keep the symmetry structure of the graph.

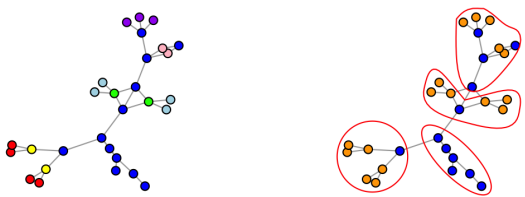


Figure 2: Input and output of the cutting algorithm.

Design based on the cutting algorithm

The procedure for obtaining a design:

1. Calculate the subgraphs using the deterministic algorithm.
2. Calculate the optimal design for each subgraph.
3. Rescale each optimal designs with the factor $\frac{I_j^{sub}}{I}$ for $j \in \{1, \ldots, k\}$.
4. Fuse the designs back together into one design for the supergraph.

Simulation study and results

Simulation study

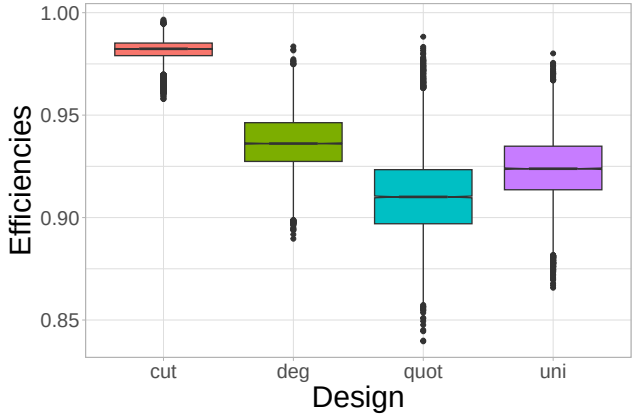


Figure 3: A-efficiencies of the different designs for 10,000 simulated networks.

Symmetries within the graph

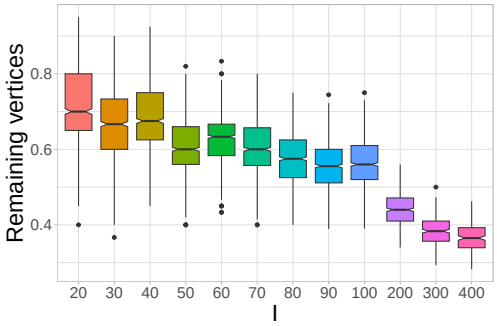


Figure 4: Relative amount of remaining vertices in the quotient graph for different network sizes.

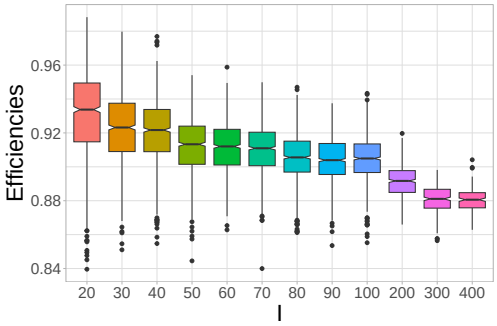


Figure 5: A-efficiencies of the quotient graph for different network sizes.

Outlook

- Designs under topological changes of the network.

One potential Approach:

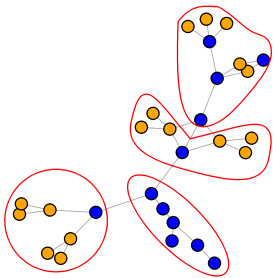


Figure 6: Cutting algorithm of the original graph.

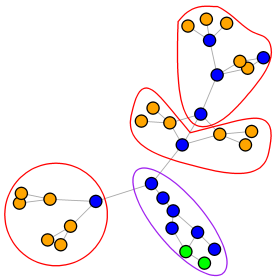


Figure 7: Cutting algorithm under topological changes.

Outlook

- Robust designs under topology changes of the network.
- Since networks in the real world might be not symmetrical, further research on the reducing of networks without using any symmetries (cutting algorithm) or other algorithms such as minimum-k cut or Kron reduction.
- Measurement devices with concrete measurement accuracies, thus transforming the underlying approximate optimization problem into a discrete optimization problem.
- Extension of the model to time-dependent errors or make the model more feasible for electrical grids e.g. using power flow equations.

References

Chung, F. R. K. (1997). Spectral Graph Theory. American Mathematical Society.

Harary, F. (1969). Graph Theory. Addison-Wesley.

Koutra, V. (2017). Designing Experiments on Networks. Phd thesis, University of Southampton.

McKay, B. D. (1981). Practical graph isomorphism. Congressus Numerantium, 30:45–87.

Müller, C. H. and Schorning, K. (2023). A-optimal designs for state estimation in networks. Statistical Papers, 64(4):1159–1186.

Stein, E. M. and Shakarchi, R. (2003). Fourier Analysis: An Introduction, volume 1 of Princeton Lectures in Analysis. Princeton University Press.

Thanks for listening