Design for Networks and Networks for Design

Ben M Parker

Brunel University, London

July 2023

Part I

Design for Networks


Ben Parker @benmparker · Sep 22, 2020
The three principles of experimental design in tweet form:
Ben Parker
@benmparker
Replication:
vs
n=2
=
n=4
=
n=large
=
11:30 AM · Sep 22, 2020
See Ben Parker's other Tweets


Ben Parker
@benmparker
Randomisation:
vs
----->
11:30 AM · Sep 22, 2020

Ben Parker
@benmparker

Blocking:
💉vs 💊

 =

 =

+ randomise
 =

11:30 AM · Sep 22, 2020

Examples of experiments we might be interested in

Different adverts are sent to different customers (experimental units). We measure the amount (£) spent on the product after advertising. This amount may be influenced by other customers in their social network.

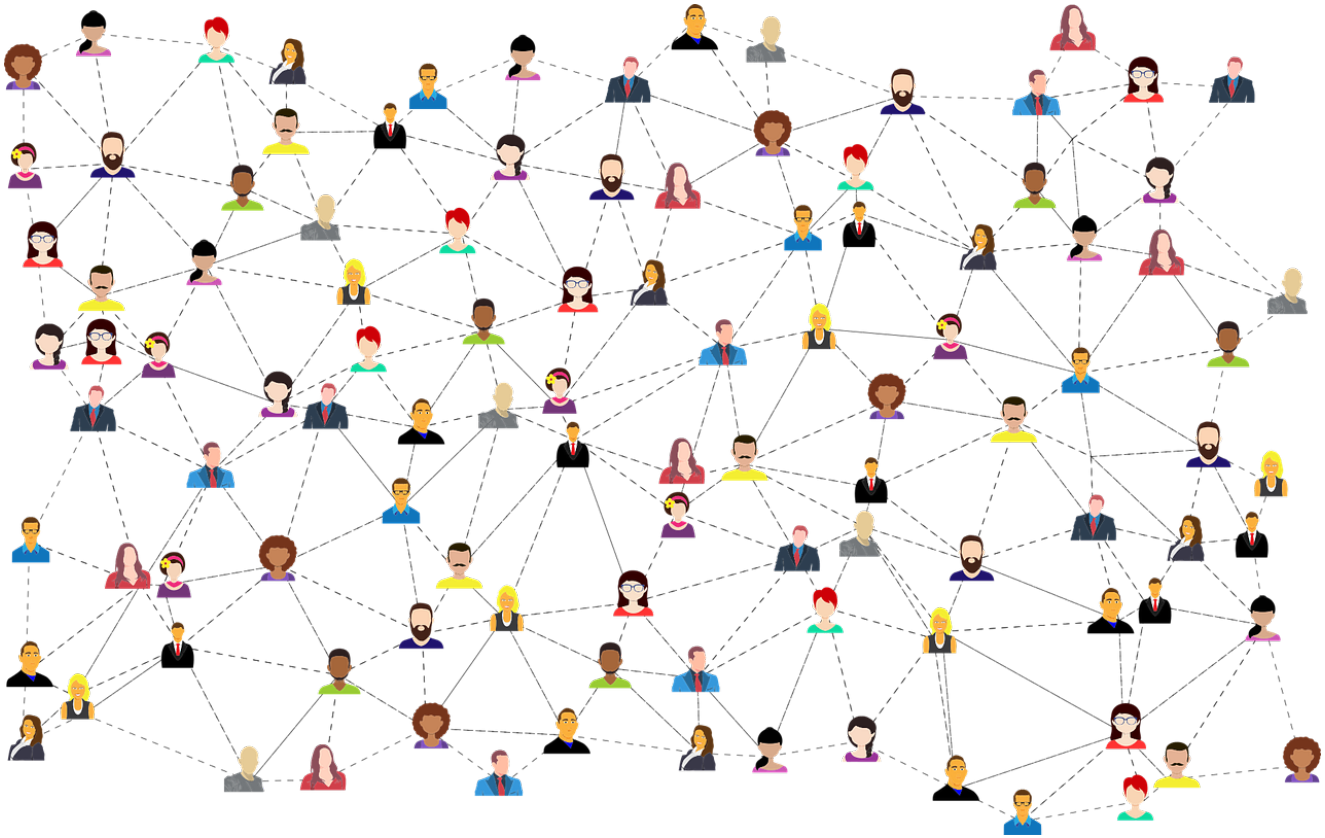


Image by Gordon Johnson from Pixabay

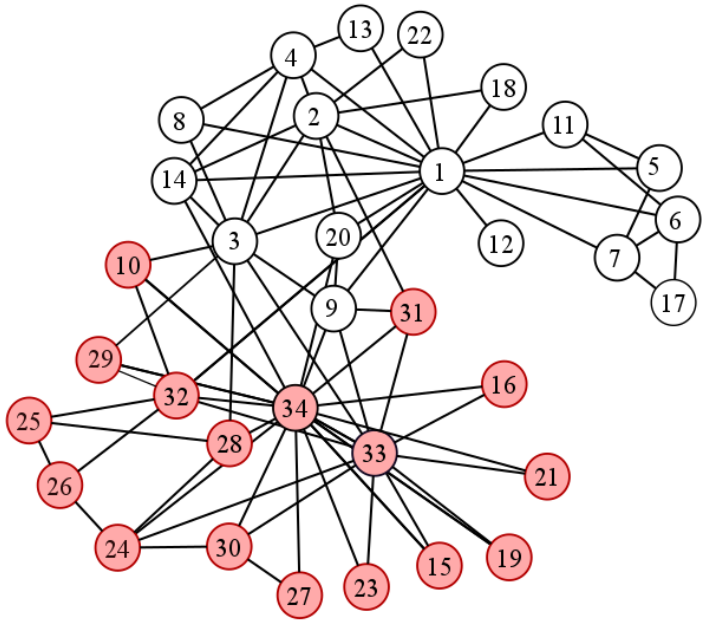
Examples of experiments we might be interested in

A crop experiment to assess the effect of fertilisers on biodiversity; we apply several treatments (fertilisers) to experimental units (plots), and measure the response (biomass of insects present) in each experimental unit after one year.

Karate club

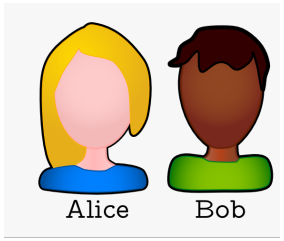
34 members of a club.

- Without considering network structure all randomly chosen designs are equivalent (17 get treatment 1, 17 get treatment 2):
  1 2 3 4 5 6 7 8 9 . . . 33 34
- If a network exists need to carefully consider best experimental design



- replication still applies
- randomisation may not be possible (not all nodes are equal)
- blocking may be a factor (groups are formed as people are different)

Networked experiments



Alice and Bob are friends. We feed Alice a new type of chocolate, and Bob a control (no chocolate).

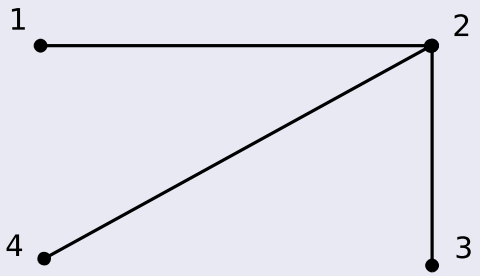
Ideas:

1. If I give a treatment to Alice, Bob may see the treatment and like it. (Indirect treatment effect: Alice has chocolate so Bob is happy!)
2. If I give a treatment to Alice, her friend Bob may see Alice's response, and modify his response.
   (Autoregressive effect: Alice is happy after eating chocolate, so Bob is happy.)
3. Alice and Bob are friends so there responses are similar (Blocking)

(1)= JRSSC Paper (BP/Gilmour/Schormans)[2]
(1+3)= Vasiliki Koutra/ BP/Gilmour [1]
(2) = Current work (discussed here)
(1+2+3)= Easy generalisation/much work to do

Networks

- We consider a graph, as a collection of nodes N and edges E.
- The nodes represent subjects on which we apply some treatment. We have $|N| = n$ subjects.
- The edges represent some relationship between the subjects.
- The relationship between our subjects is specified by the adjacency matrix A where $A_{ij} > 0$ if i and j are related and $A_{ij} = 0$ otherwise. By convention, $A_{ii} = 0$.
- We sometimes assume additionally that links are non-directional, such that $A_{ij} = A_{ji}$, i.e. A is symmetrical.



$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \text{ or } A = \begin{pmatrix} 0 & 0.5 & 0 & 0 \\ 0.3 & 0 & 0.2 & 0.1 \\ 0 & 0.4 & 0 & 0 \\ 0 & 0.4 & 0 & 0 \end{pmatrix}$$

Spatial Lag model

The response of each node depends on the response of connected units:

$$Y_i = \mu + \tau_{t(i)} + \rho \sum_{k=1}^{n} A_{ik} Y_k + \epsilon_i, \tag{1}$$

where

- $i \in \{1, \ldots, n\}$ is the experimental unit,
- $t(i) \in \{1, \ldots, m\}$ is treatment given to the i-th experimental unit,
- τ_j is the treatment effect of the $j-$th treatment, $j \in \{1, \ldots, n\}$,
- ρ is an autoregressive parameter,
- A is a $n \times n$ matrix where $0 \leq A_{ij} \leq 1 \quad \forall i \neq j$, $A_{ii} = 0 \quad \forall i$,
- we assume that ϵ_i are independent and identically normally distributed with mean 0 and constant variance σ^2.

Each experimental unit receives exactly one (unstructured) treatment.
WLOG, to allow identifiability of the treatment effects we restrict $\tau_m = 0$.

Let $K = K(\rho) = (I - \rho A^T)^{-1}$.
We can show

$$Y \sim N(KX\beta, \sigma^2 KK^T)$$

After some algebra, we can show that the Fisher information matrix is $I =$

$$\frac{1}{\sigma^2}\begin{pmatrix} n & n_1 & n_2 & \dots & n_m & \mathbf{1}^T AKX\beta \\ n_1 & n_1 & 0 & \dots & 0 & (X_{11} \quad X_{21} \quad \dots \quad X_{n1})AKX\beta \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ n_j & 0 & \dots & \dots & 0 & (X_{1j} \quad X_{2j} \quad \dots \quad X_{nj})AKX\beta \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ n_m & 0 & 0 & \dots & n_m & (X_{1m} \quad X_{2m} \quad \dots \quad X_{nm})AKX\beta \\ \mathbf{1}^T AKX\beta & \dots & \dots & \dots & \dots & (AKX\beta)^T(AKX\beta) + \frac{\sigma^2}{2}\,\mathrm{tr}\left([K^T A^T + AK]^2\right) \end{pmatrix}$$

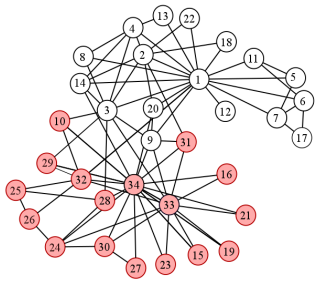
for $\theta = (\mu \quad \tau_1 \quad \tau_2 \dots \tau_m \quad \rho)^T$,

Optimality Criteria

We may be interested in estimating τ, ρ, or sometimes both.

- Estimating ρ We wish to estimate ρ as precisely as possible, which corresponds to minimising $\operatorname{Var} \hat{\rho}$.
- Estimating τ We wish to estimate the treatment effect difference as precisely as possible, so seek to minimise $\operatorname{Var} \widehat{\tau_1 - \tau_2}$.

These both correspond to maximising some function of the information matrix, I.



$\rightarrow$ Assign treatments (label nodes with 1s and 2s) $\rightarrow$ Check f(I) maximised.

Results

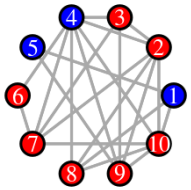
As Fisher Information depends on unknown parameters β, ρ, σ^2, we arbitrarily chose Locally Optimal Designs:, i.e. optimal for a particular value of these parameters that in reality may be unknown before doing the experiment.

This is equivalent to choosing a point prior for the unknown parameters. We chose $\beta_0 = \mu = 0, \beta_1 = \tau_1 = 10, \beta_2 = \tau_2 = 20, \sigma^2 = 1$ and present results for $\rho = 0, 0.1, 0.2, \ldots, 0.9$.

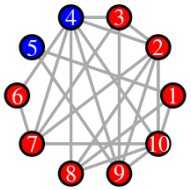
We demonstrate for $m = 2$ treatments. Treatment 1 is shown in red, treatment 2 in blue.

Optimal designs for Estimating ρ

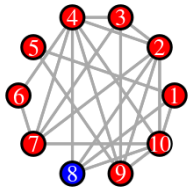
rho = 0



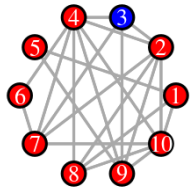
rho = 0.1



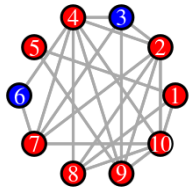
rho = 0.2



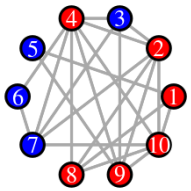
rho = 0.3



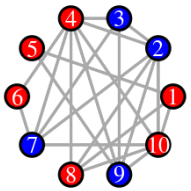
rho = 0.4



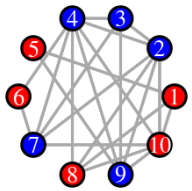
rho = 0.5



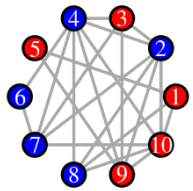
rho = 0.6



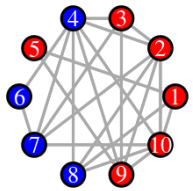
rho = 0.7



rho = 0.8

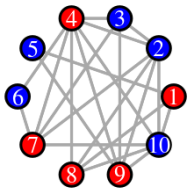


rho = 0.9

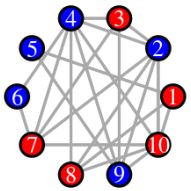


Optimal designs for Estimating τ

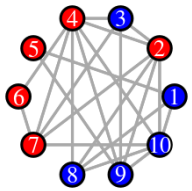
rho = 0



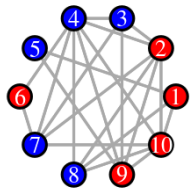
rho = 0.1



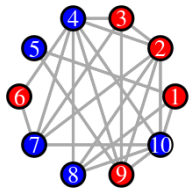
rho = 0.2



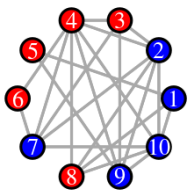
rho = 0.3



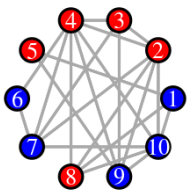
rho = 0.4



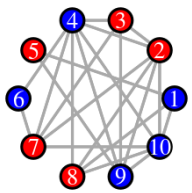
rho = 0.5



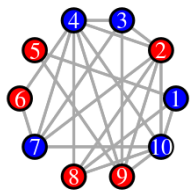
rho = 0.6



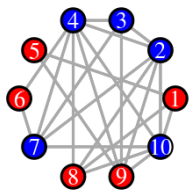
rho = 0.7



rho = 0.8

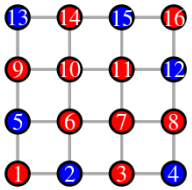


rho = 0.9

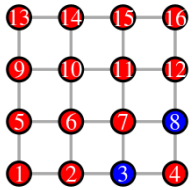


Optimal designs for Estimating ρ

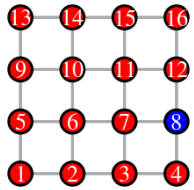
rho = 0



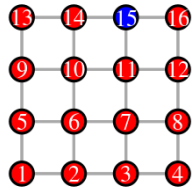
rho = 0.1



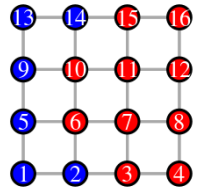
rho = 0.2



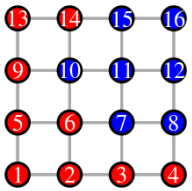
rho = 0.3



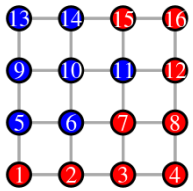
rho = 0.4



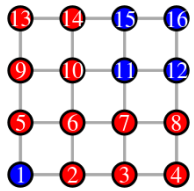
rho = 0.5



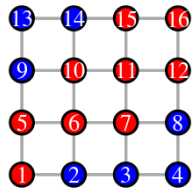
rho = 0.6



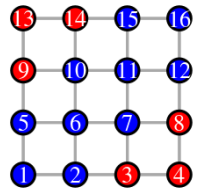
rho = 0.7



rho = 0.8

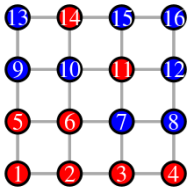


rho = 0.9

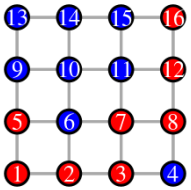


Optimal designs for Estimating τ

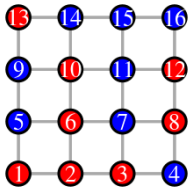
rho = 0



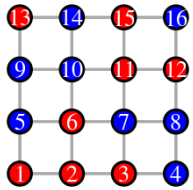
rho = 0.1



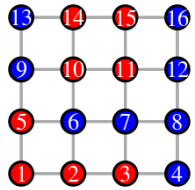
rho = 0.2



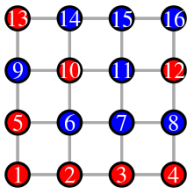
rho = 0.3



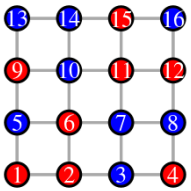
rho = 0.4



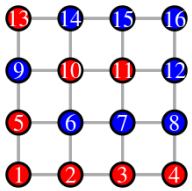
rho = 0.5



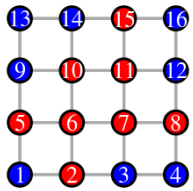
rho = 0.6



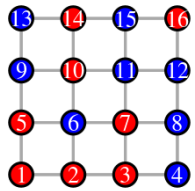
rho = 0.7



rho = 0.8



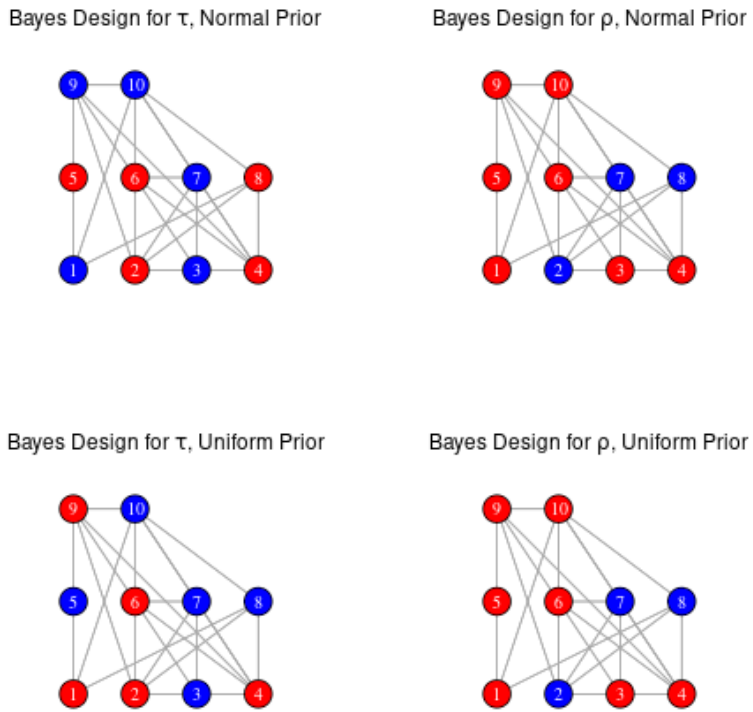
rho = 0.9



Bayesian Priors

For a simple network, specify two simple priors:

- (Top) Prior 1: $\mu = 2$, $\tau_2 = 1$, $\tau_2 = 0$, and $\rho \sim N(0.5, \sigma^2), |\rho| < 1$.
- (Left) Prior 2: As prior 1, except now the prior was $\rho \sim U[0, 1]$



(pseudo)-Bayesian designs for two simple priors on ρ, under the MRSAR for estimating (left)ψ_τ and (right) ψ_ρ

Efficiency

We chose point priors of
$\beta_0 = \mu = 0 = 20, \beta_1 = \tau_1 = 10, \beta_2 = \tau_2 = 0, \sigma^2 = 1$, and now $\rho = 0.5$.

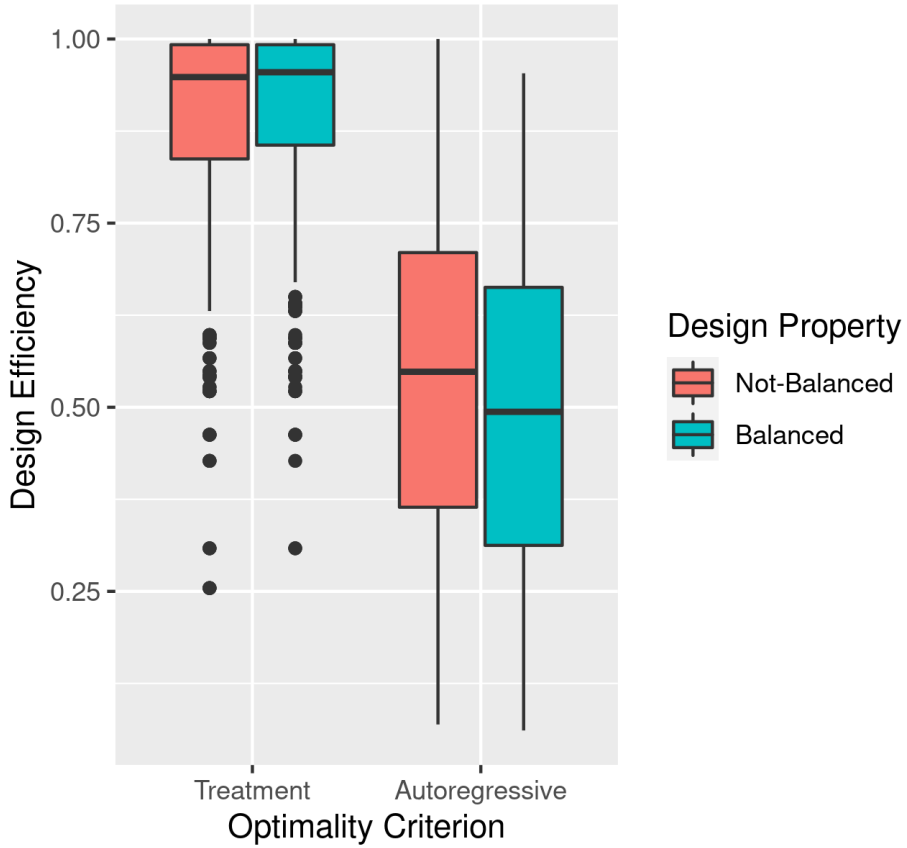


Figure: Efficiency of optimal designs. Left: efficiency for ψ_τ, right hand pair of plots the efficiency forψ_ρ.

Not just for simple unweighted network models...

Up to now, examples have used adjacency matrices A which are binary, but nothing in our theory requires this to be so.

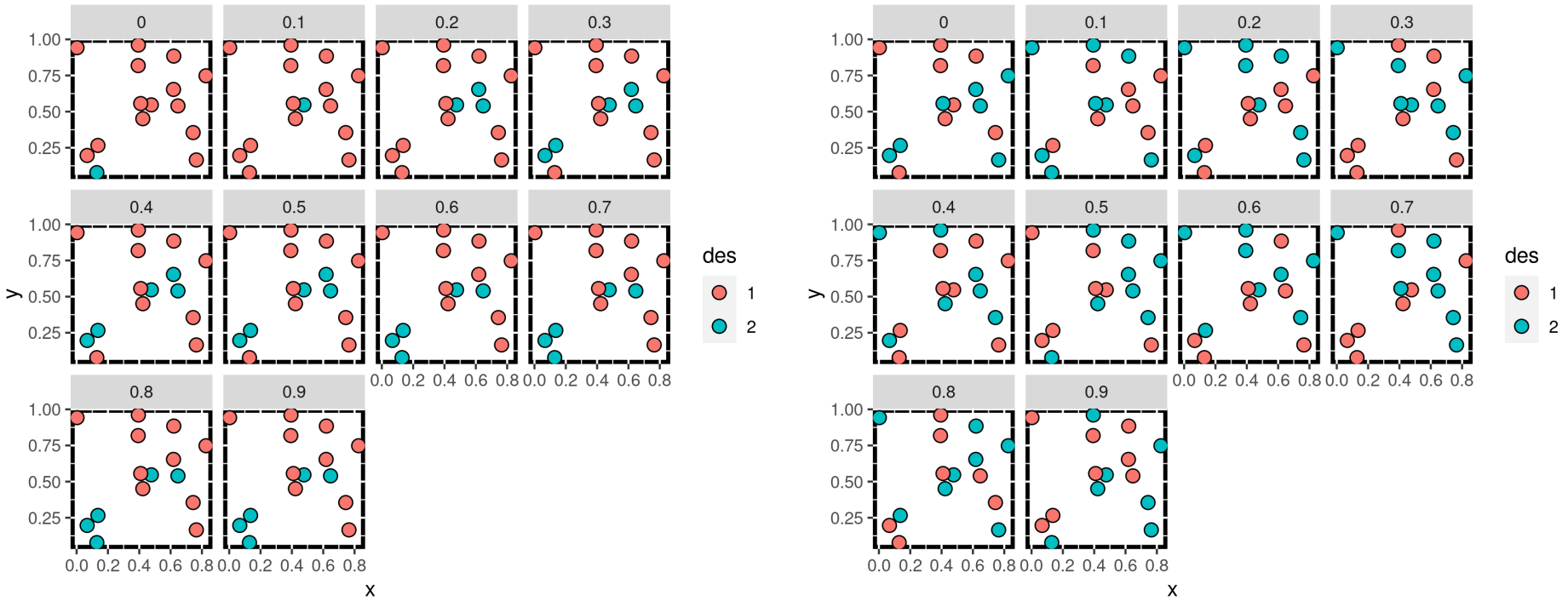


Figure: Optimal spatial designs for points randomly chosen in $[0, 1] \times [0, 1]$ for varying ρ. (Left) shows optimal designs for ψ_ρ, estimating AR parameter. (right) shows optimal designs for ψ_τ, estimating treatment difference $\tau_2 - \tau_1$.

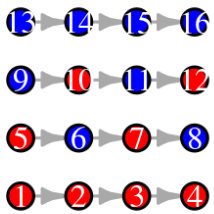
Crossover trial

In a crossover trial, generally each subject/period combination is an experimental unit.

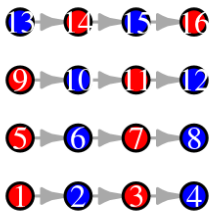
		Period			
		1	2	3	4
Subject	a	1	2	3	4
	b	5	6	7	8
	c	9	10	11	12
	d	13	14	15	16

Optimal designs for Estimating τ

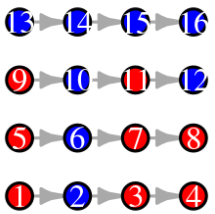
rho = 0



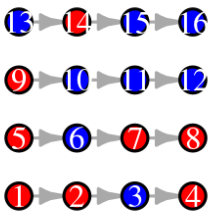
rho = 0.1



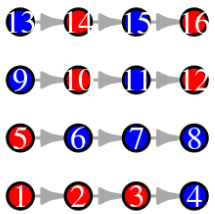
rho = 0.2



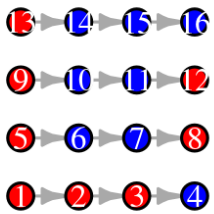
rho = 0.3



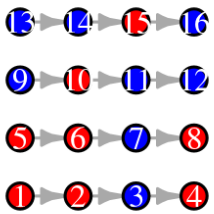
rho = 0.4



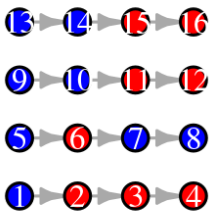
rho = 0.5



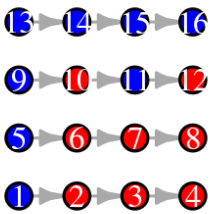
rho = 0.6



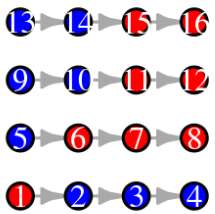
rho = 0.7



rho = 0.8



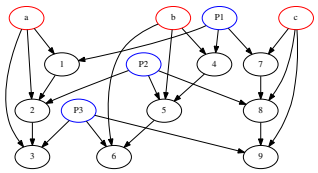
rho = 0.9



Part II

Networks for Design

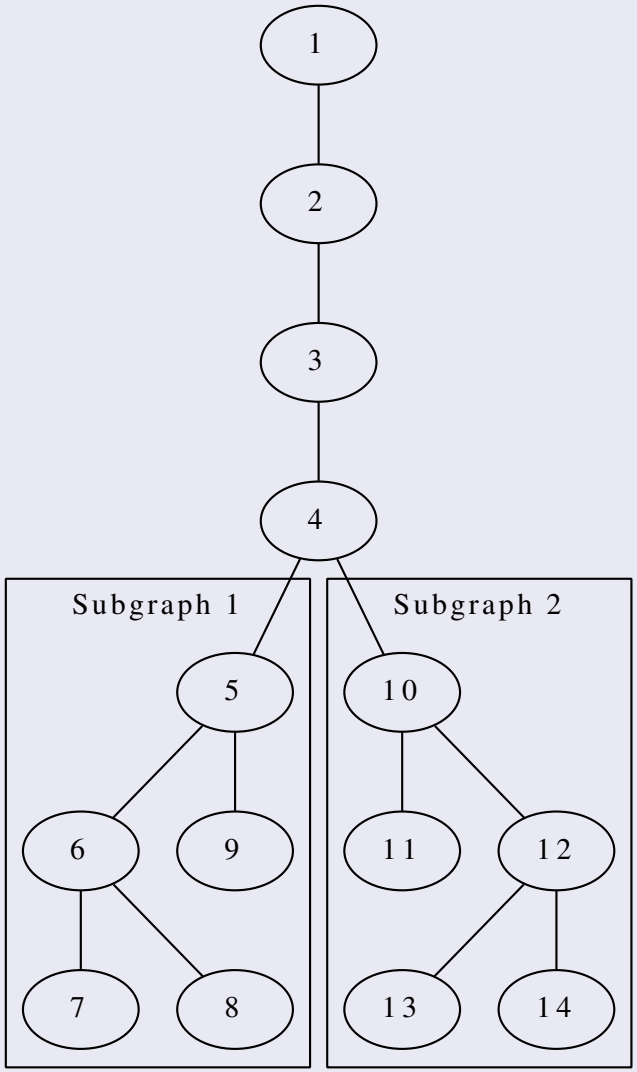
Design Algorithm



$\rightarrow$ Assign treatments (label nodes with 1s and 2s) $\rightarrow$ Check f(I) maximised.

- The design space is all possible colourings of the graph of n nodes with m colours representing m treatments.
- We can use a complete enumeration, or well known algorithms such as exchange algorithms.
- For even reasonable networks the design space (the set of all possible colourings) is very large.

Symmetry of graphs



Subgraphs 1 and 2 are exchangeable; i.e for subdesign A on subgraph 1, and subdesign B on subgraph 2 (call this [A1,B2]), we need not also consider [A2,B1] as by symmetry this design has the same criterion value.

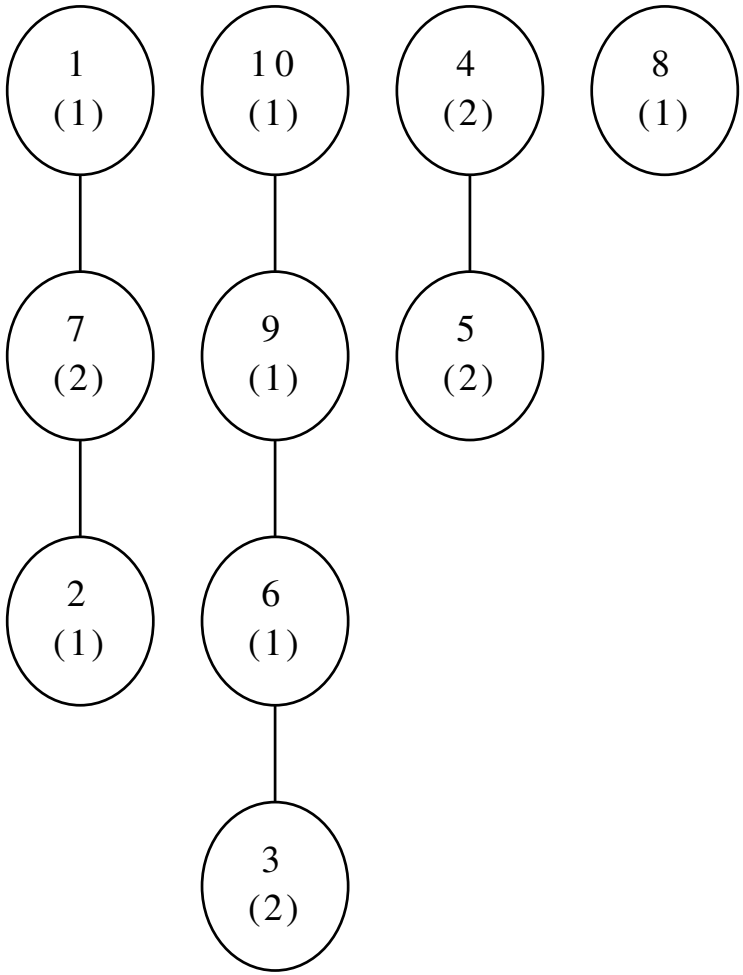
We can reduce our design space greatly if we can identify exchange subgraphs.

This is equivalent to finding an automorphism for our network, a relabelling or permutation of the set N such that the edges E are preserved.

This is the Graph Automorphism Problem.

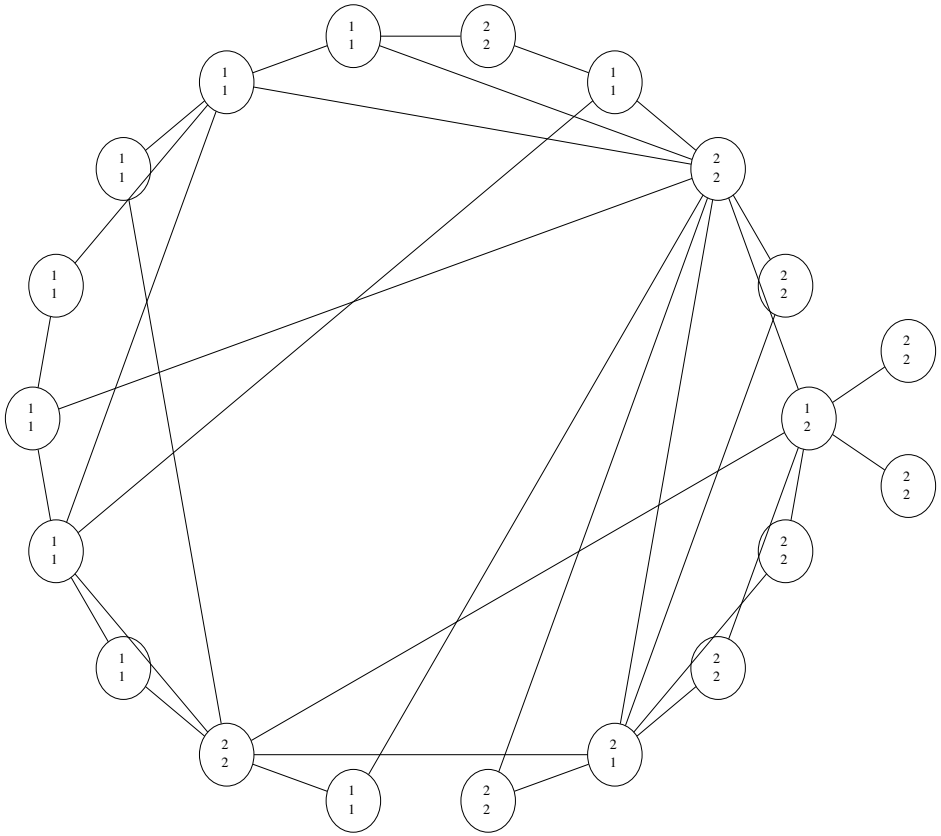
Recently it has been shown (and peer reviewed?) that this is a hard problem, but is not (quite) NP-hard. However, fast algorithms do exist in software, including R.

Example 1



We obtain the correct optimal design, but evaluate the information matrix 236 times as opposed to 507 times.

Example 2

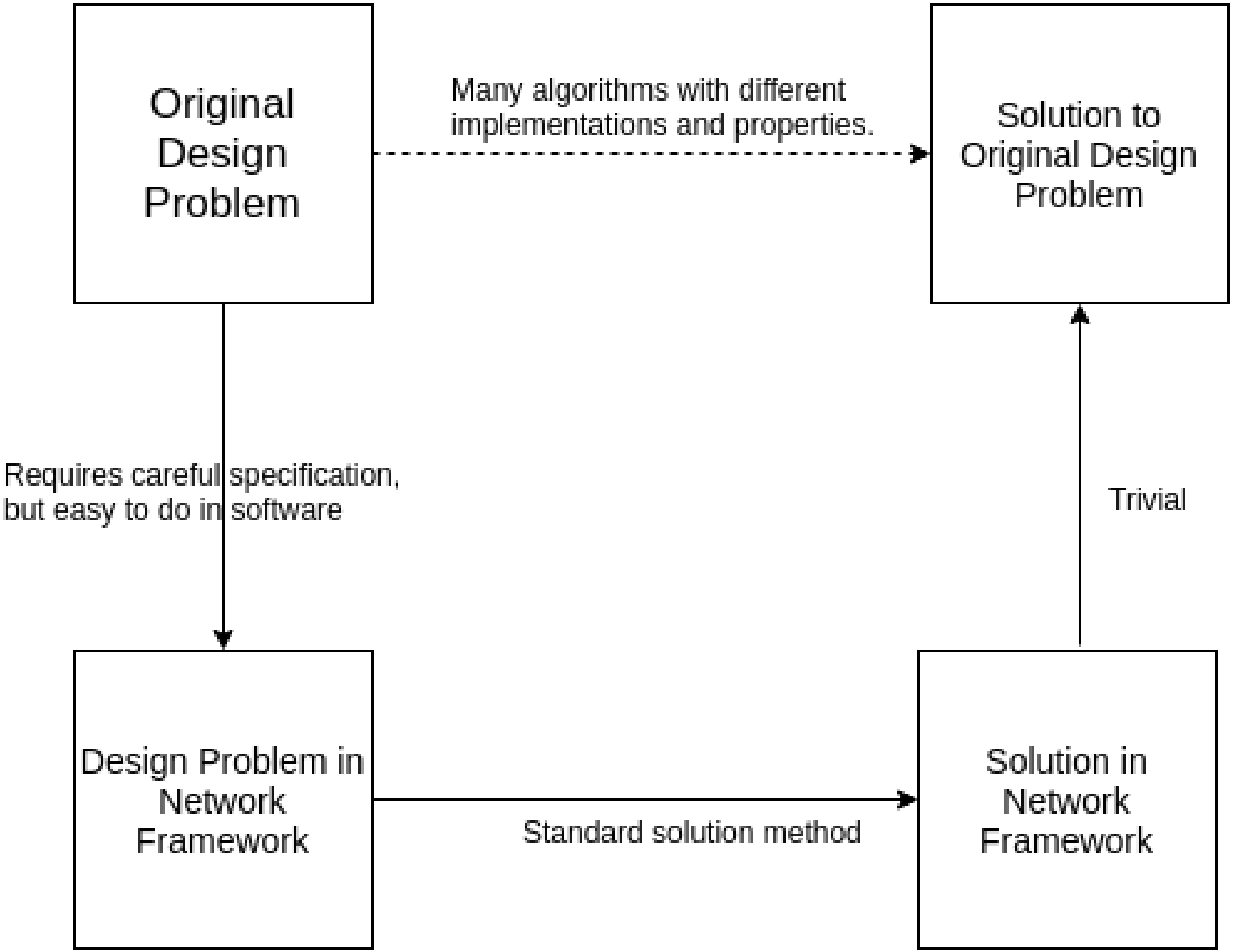


524287 without automorphisms, 221183 with.
58.58s without, opposed to 31.56 with.

Results

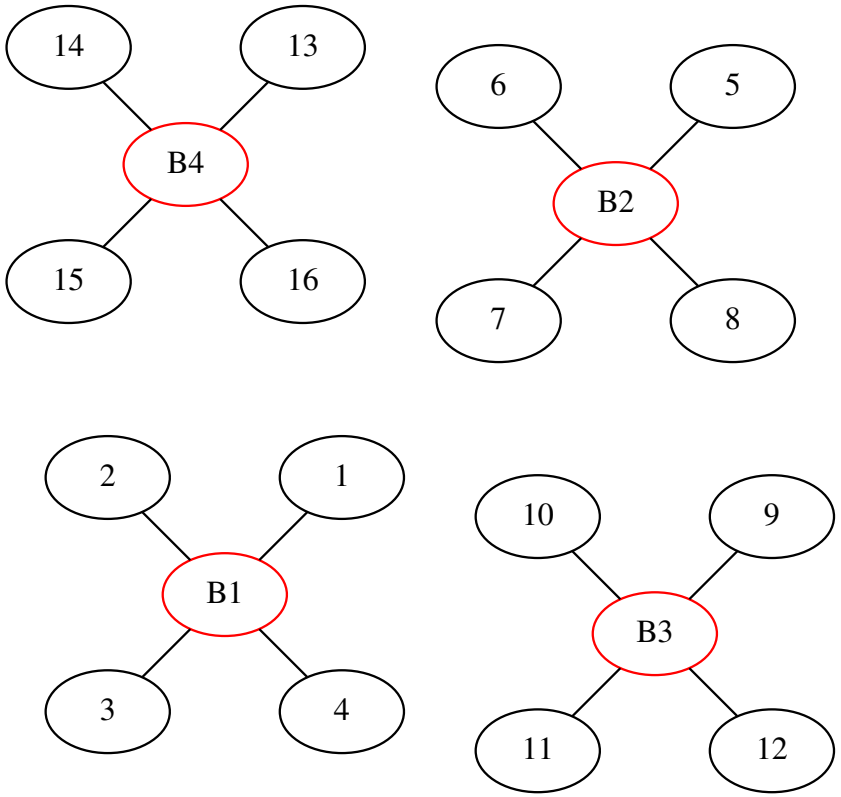
Description	n	No. auto-morphisms	Evaluations without auto-morphisms	Evaluations with auto-morphisms
Small social network (1)	10	8	507	236
Small social network	10	1	511	511
Larger social network (2)	20	8	524,287	221,183
Block design with neighbour effects	12	384	535,008	18,766
Non-rectangular field trial (3)	15	2	2,368,741	1,581,572
Crossover trial with dropouts(4)	15	6	2,262,800	904,555

Description	n	No. automorphisms	Time without automorphisms	Time with automorphisms
Small social network (1)	10	8	0.04	0.02
Small social network	10	1	0.04	0.04
Larger social network (2)	20	8	58.58	31.56
Block design with neighbour effects	12	384	108.52	33.68
Non-rectangular field trial (3)	15	2	279.6	197.58
Crossover trial with dropouts(4)	15	6	283.86	134.26

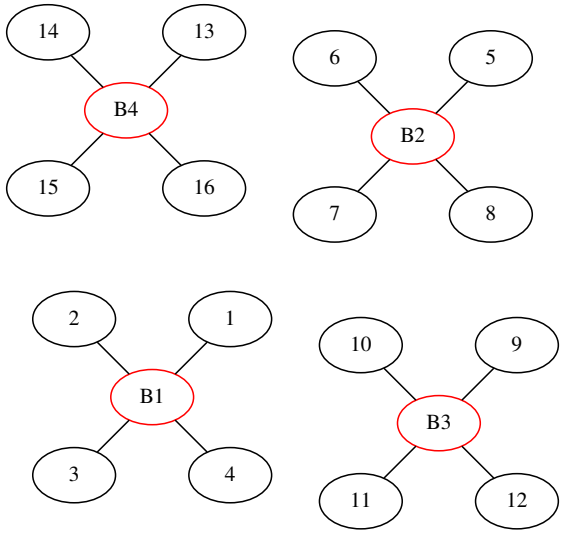

Original Design Problem
Many algorithms with different implementations and properties.
Solution to Original Design Problem
Requires careful specification, but easy to do in software
Trivial
Design Problem in Network Framework
Standard solution method
Solution in Network Framework

Simple Blocked Experiment

Block	Experimental Units			
1	1	2	3	4
2	5	6	7	8
3	9	10	11	12
4	13	14	15	16



Block	Exp. Units			
1	1	2	3	4
2	5	6	7	8
3	9	10	11	12
4	13	14	15	16



	Original Problem	Network Problem
No of Treatments	2	6
Exp. Units	$\{1, 2, \ldots 16\}$	$\{1, 2, \ldots, 16, B1, B2, B3, B4\}$
Wish to estimate	$\tau_1 - \tau_2$	$\tau_1 - \tau_2$
Opt. criterion	A	A_s
Restrictions	Can apply either treatment to any unit.	Can apply treatments 1,2 to units $\{1, 2, \ldots, 16\}$, and treatments 3,4,5,6 to units B1-B4.

Block designs are network designs!

We represent our blocked experiment by the linear network effects model as

$$Y_i = \mu + \tau_{t(i)} + \sum_{k=\{1,\ldots,16,B1,B2,B3,B4\}} A_{ik}\gamma_{t(k)} + \epsilon_i, \quad i = 1, \ldots, 16$$

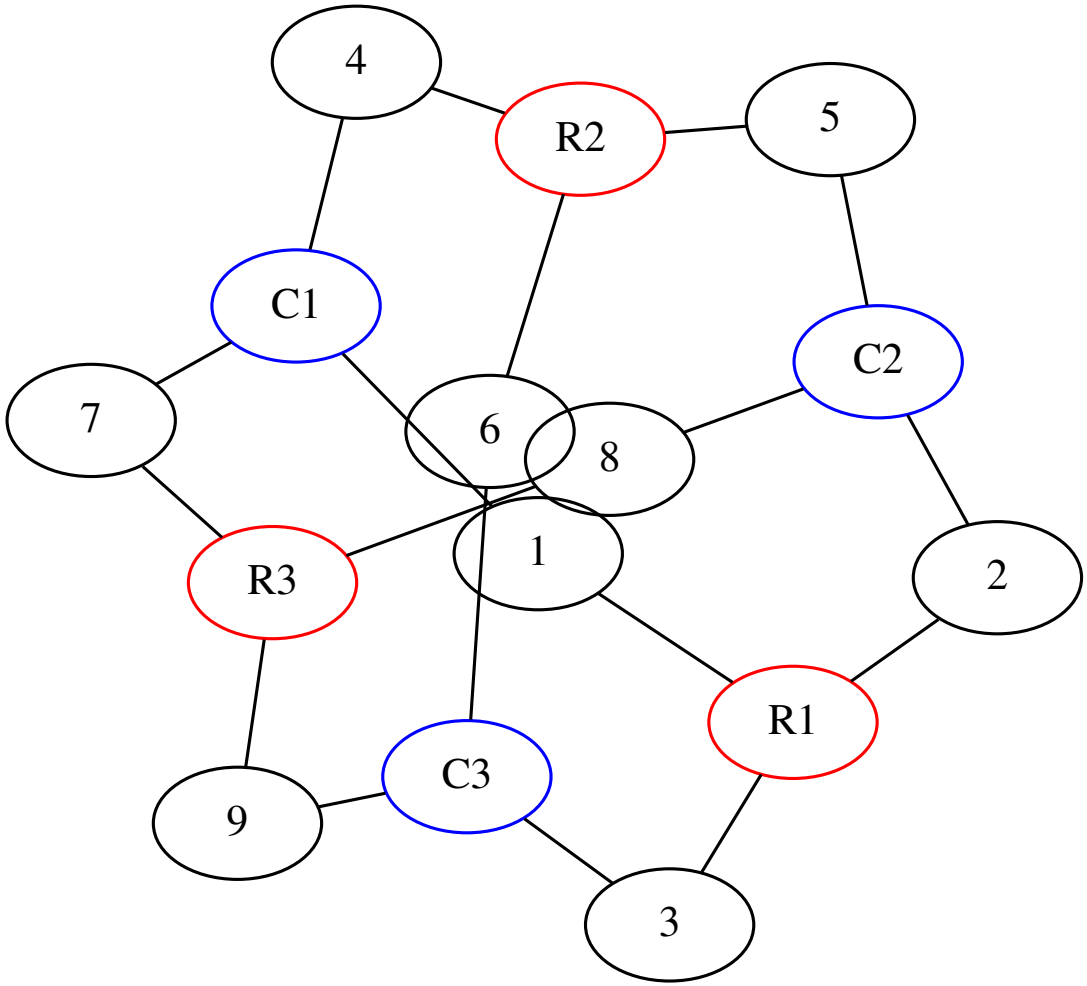
By writing $b_{j(i)} = \sum_{k=\{1,\ldots,16,B1,B2,B3,B4\}} A_{ik}\gamma_{t(k)}$, we can show that this is equivalent to

$$Y_i = \mu + \tau_{t(i)} + b_{j(i)} + \epsilon_i,$$

a more familiar representation of a blocked experiment, where $b_{j(i)}$ is block effect of experimental unit i being in block j.

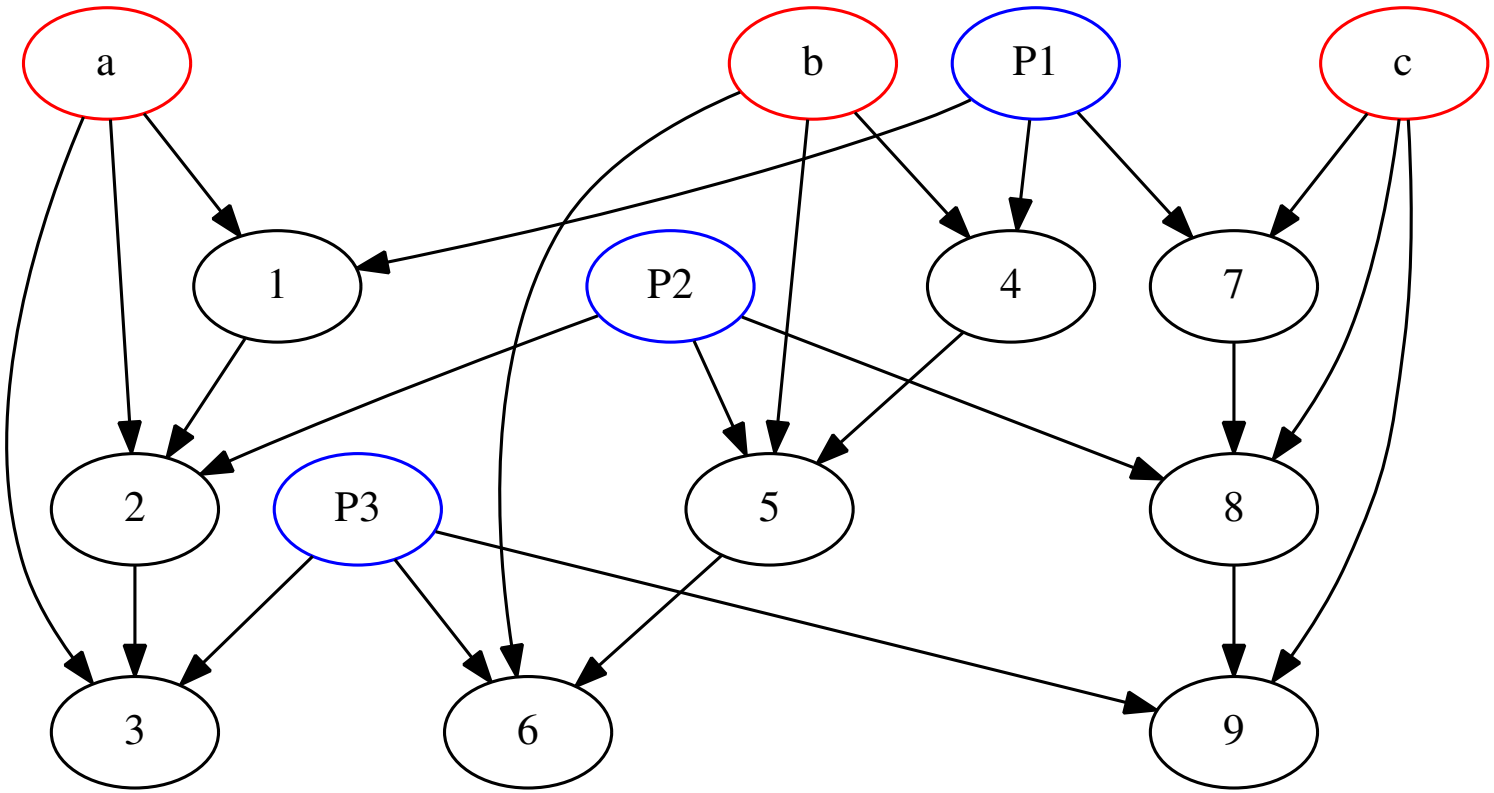
Row Column Design

	C1	C2	C3
R1	1	2	3
R2	4	5	6
R3	7	8	9



Crossover Design

		Period		
		P1	P2	P3
Subject	a	1	2	3
	b	4	5	6
	c	7	8	9



Evaluating blocked designs

We evaluate several blocked experimental structures:

1. 3 blocks of size 3 ($n = 9$), with 3 treatments. The optimal designs are randomised complete block designs;
2. 4 blocks of size 3 ($n = 12$), with i) 3 and ii) 4 treatments. The optimal designs are i) randomised complete block designs and ii) balanced incomplete block designs;
3. A row-column structure with 3 rows and 3 columns, each row-column intersection containing a single experimental unit, with 3 treatments.

   The optimal designs are Latin Squares of size 3:

1	2	3
2	3	1
3	1	2

4. A row-column structure with 4 rows and 4 columns, each row-column intersection containing a single experimental unit, with i)3 and ii)4 treatments. The optimal designs for ii) are Latin Squares of size 4.

Results for blocked designs

Ex.	Description	n	m	Number of automorphisms	Evaluations without automorphisms	Evaluations with automorphisms
1	3x3 Blocks	9	3	1,296	2,925	94
2i	4x3 Blocks	12	3	82,944	86,126	379
2ii	4x3 Blocks	12	4	82,944	605,960	1808
3	3x3 Row Column	9	3	72	2,807	241
4i	4x4 Row Column	16	3	1,152	7,123,656	34,873
4ii	4x4 Row Column	16	4	1,152	170,863,644	1,610,909

Results for blocked designs

Ex.	Description	n	m	Number of automorphisms	Time without automorphisms	Time with automorphisms
1	3x3 Blocks	9	3	1,296	2.52	1.54
2i	4x3 Blocks	12	3	82,944	55.44	310.02
2ii	4x3 Blocks	12	4	82,944	378.82	1,051.54
3	3x3 Row Column	9	3	72	1.9	0.48
4i	4x4 Row Column	16	3	1,152	6,051.12	493.32
4ii	4x4 Row Column	16	4	1,152	141,456.6	14,123.94

We don't claim that these designs would be sensibly found via this method, as the solutions are known analytically, but we seek to demonstrate the benefits of automorphisms via improvements in calculations.

Experimental design principles

Remember the three Rs?

- ~~Replication~~
- Randomisation = Automorphisms of graph
- ~~r...Blocking~~

Conclusions and future work

- These network models allows a wide range of experiments (networks/spatial/temporal) to be considered, even when a network is not obvious. Suggestions for real experiments in the literature would be very useful.
- R Software is available at `https://github.com/bmp22/networkDesign`. The documentation for spatial lag model is still in progress, but there are some vignettes/instructions for simpler models
- Still some challenges on analysis of data from experiments designed in this way.
- Some computational challenges to find exact designs for a wide range of large networks, but finding good designs which are highly efficient is very possible.

References


Vasiliki Koutra, Steven G Gilmour, and Ben M Parker.
Optimal block designs for experiments on networks.
Journal of the Royal Statistical Society: Series C (Applied Statistics), 2021.


Ben M. Parker, Steven G. Gilmour, and John Schormans.
Optimal design of experiments on connected units with application to social networks.
Journal of the Royal Statistical Society: Series C (Applied Statistics), 66(3):455–480, April 2016.

Extending the model

A more general model is

$$Y_{ij} = \mu + \tau_{t(i)} + b_j + \sum_{k=1}^{n} A_{ki}\gamma_{t(k)} + \rho \sum_{k=1}^{n} A_{ik} Y_k + \epsilon_{ij}, \tag{2}$$

where the response of the (ij)-th experimental unit now depends on

- a grand mean μ
- a treatment effect corresponding to the treatment given directly to the experimental unit;
- a block effect depending on the block that the experimental unit is in;
- the sum of the network effect corresponding to treatments given to neighbours of the experimental unit, weighted for the strength of the influences of those neighbours;
- the sum of the "viral" effects due to the experimental units response being modified due to the response of its neighbours;
- an error term.

This model in vector form is

$$E(Y) = X(\mu \quad \tau^T)^T + Z\mathbf{b} + A^T X\gamma + \rho A^T \mathbf{Y} + \epsilon$$

where Z is the block incidence matrix such that $Z_{ij} = 1$ if experimental unit i is in block j, and $Z_{ij} = 0$ otherwise, and all other terms are as before. We define our vector of unknown parameters β here as

$$\beta = (\mu \quad \tau^T \quad \mathbf{b}^T \quad \gamma^T)^T,$$

and show similarly to before that our information matrix is now given by the block matrix

$$I = \frac{1}{\sigma^2} \left(\begin{array}{c|c|c|c} X^T X & X^T Z & X^T AX & X^T AKX\beta \\ \hline & Z^T Z & Z^T AX & Z^T AKX\beta \\ \hline & & (AX)^T AX & (AX)^T AKX\beta \\ \hline & & & (AKX\beta)^T (AKX\beta) + \frac{\sigma^2}{2} \operatorname{tr}\left([K^T A^T + AK]^2\right) \end{array} \right)$$

where only the upper diagonal terms are shown as the matrix is symmetric.

We also plot, for the same network, the information criterion for 3 arbitrary balanced designs for $0 < \rho < 1$, with prior for other variables as before.

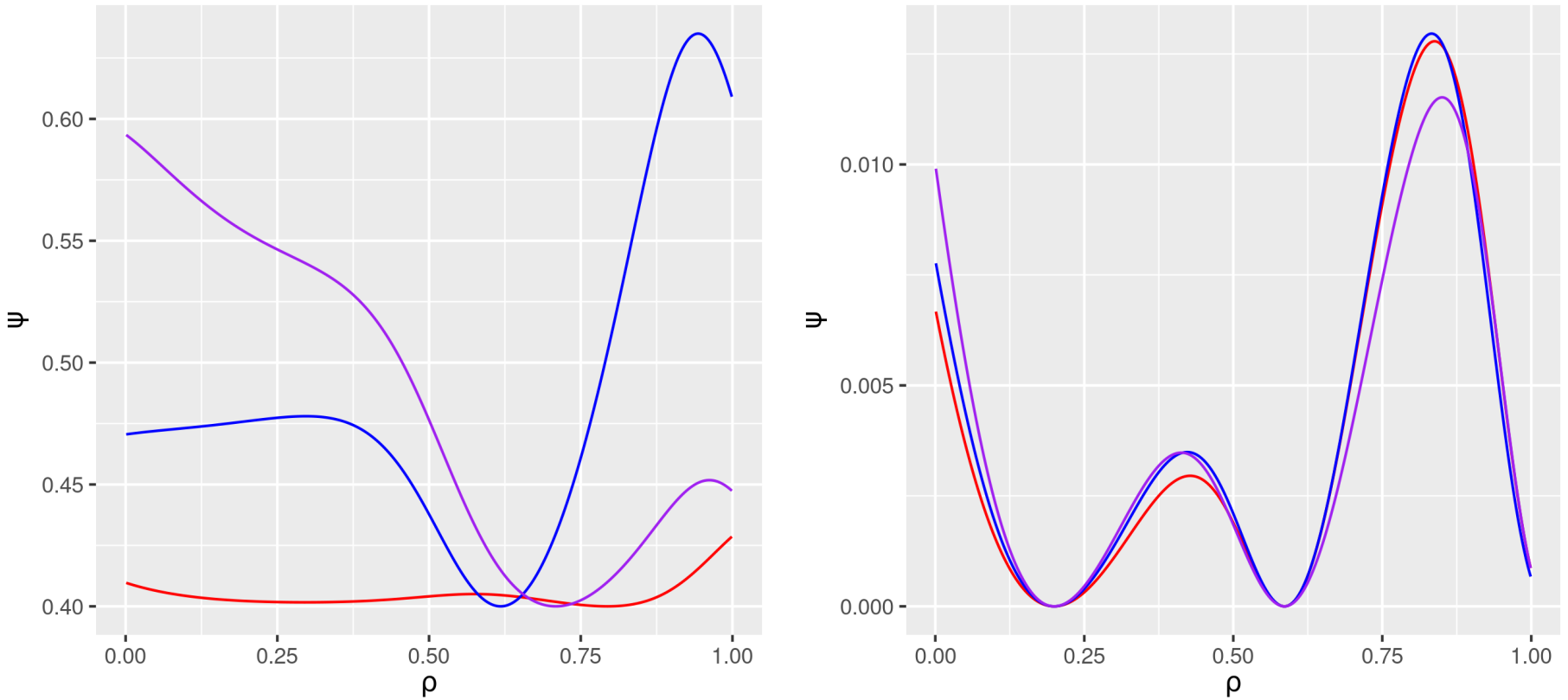


Figure: Optimality function values for (left)ψ_τ and (right)ψ_ρ for varying ρ for two arbitrary balanced designs: (1,1,1,1,1,2,2,2,2,2) in red, (1,2,1,2,1,2,1,2,1,2) in blue, and (1,1,2,2,1,1,2,2,1,2) in purple

This shows clearly that ψ_τ varies substantially for this network depending on the (usually unknown) value of ρ, but that ψ_ρ varies less.