Block designs on the edge

R. A. Bailey

Queen Mary University of London

r.a.bailey@qmul.ac.uk

Finite Geometries Third Irsee Conference: June 2011

What makes a block design good for experiments?

I have v treatments that I want to compare.
I have b blocks,
with space for k treatments (not necessarily distinct) in each block.
How should I choose a block design?

Two designs with $v = 15$, $b = 7$, $k = 3$: which is better?

1	1	2	3	4	5	6
2	4	5	6	10	11	12
3	7	8	9	13	14	15

replications differ by ≤ 1

1	1	1	1	1	1	1
2	4	6	8	10	12	14
3	5	7	9	11	13	15

queen-bee design

The replication of a treatment is its number of occurrences.

A design is a queen-bee design if there is a treatment that occurs in every block.

Average replication $= \bar{r} = bk/v = 1.4$.

Two designs with $v = 5$, $b = 7$, $k = 3$: which is better?

1	1	1	1	2	2	2
2	3	3	4	3	3	4
3	4	5	5	4	5	5

binary

1	1	1	1	2	2	2
1	3	3	4	3	3	4
2	4	5	5	4	5	5

non-binary

A design is binary is no treatment occurs more than once in any block.

Average replication $= \bar{r} = bk/v = 4.2$.

Two designs with $v = 7$, $b = 7$, $k = 3$: which is better?

1	2	3	4	5	6	7
2	3	4	5	6	7	1
4	5	6	7	1	2	3

balanced (2-design)

1	2	3	4	5	6	7
2	3	4	5	6	7	1
3	4	5	6	7	1	2

non-balanced

A binary design is balanced if every pair of distinct treatments occurs together in the same number of blocks.

Average replication $=$ every replication $= \bar{r} = bk/v = 3$.

Design $\longrightarrow$ graph

If $i \neq j$, the concurrence λ_{ij} of treatments i and j is the number of occurrences of the pair $\{i, j\}$ in blocks, counted according to multiplicity.

Design $\longrightarrow$ graph

If $i \neq j$, the concurrence λ_{ij} of treatments i and j is
the number of occurrences of the pair $\{i, j\}$ in blocks,
counted according to multiplicity.

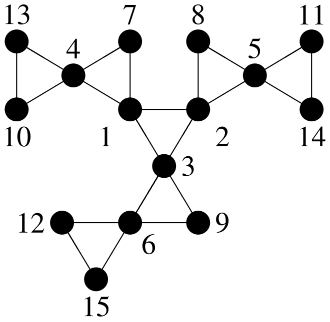
The concurrence graph G of the design has the treatments as vertices.
There are no loops.
If $i \neq j$ then there are λ_{ij} edges between i and j.
So the valency d_i of vertex i is

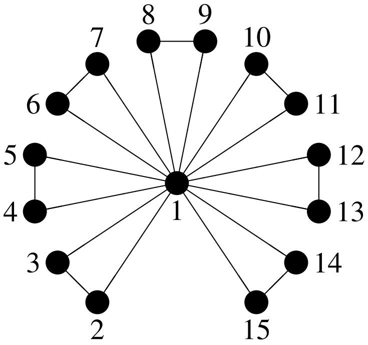
$$d_i = \sum_{j \neq i} \lambda_{ij}.$$

Concurrence graphs of two designs: $v = 15$, $b = 7$, $k = 3$

1	1	2	3	4	5	6
2	4	5	6	10	11	12
3	7	8	9	13	14	15

1	1	1	1	1	1	1
2	4	6	8	10	12	14
3	5	7	9	11	13	15





Graph $\rightarrow$ matrix

The Laplacian matrix L of this graph has
(i,i)-entry equal to $d_i = \sum_{j\neq i} \lambda_{ij}$
(i,j)-entry equal to $-\lambda_{ij}$ if $i \neq j$.
So the row sums of L are all zero.

Graph $\rightarrow$ matrix

The Laplacian matrix L of this graph has

(i,i)-entry equal to $d_i = \sum_{j \neq i} \lambda_{ij}$

(i,j)-entry equal to $-\lambda_{ij}$ if $i \neq j$.

So the row sums of L are all zero.

Hence L has eigenvalue 0 on the all-1 vector.

Graph $\longrightarrow$ matrix

The Laplacian matrix L of this graph has

(i,i)-entry equal to $d_i = \sum_{j \neq i} \lambda_{ij}$

(i,j)-entry equal to $-\lambda_{ij}$ if $i \neq j$.

So the row sums of L are all zero.

Hence L has eigenvalue 0 on the all-1 vector.

This *trivial* eigenvalue has multiplicity 1

$\iff$ the graph G is connected

$\iff$ all contrasts between treatment parameters are estimable.

Graph $\longrightarrow$ matrix

The Laplacian matrix L of this graph has

(i,i)-entry equal to $d_i = \sum_{j \neq i} \lambda_{ij}$

(i,j)-entry equal to $-\lambda_{ij}$ if $i \neq j$.

So the row sums of L are all zero.

Hence L has eigenvalue 0 on the all-1 vector.

This *trivial* eigenvalue has multiplicity 1

$\iff$ the graph G is connected

$\iff$ all contrasts between treatment parameters are estimable.

Call the remaining eigenvalues *nontrivial*. They are all non-negative.

Estimation and variance

We measure the response Y on each unit in each block.

If that unit has treatment i and block m, then we assume that

$$Y = \tau_i + \beta_m + \text{random noise}.$$

We want to estimate contrasts $\sum_i x_i \tau_i$ with $\sum_i x_i = 0$.

In particular, we want to estimate all the simple differences $\tau_i - \tau_j$.

Put $V_{ij} =$ variance of the best linear unbiased estimator for $\tau_i - \tau_j$.

We want all the V_{ij} to be small.

How do we calculate variance?

Theorem

Assume that all the noise is independent, with variance σ^2.
If $\sum_i x_i = 0$, *then the variance of the best linear unbiased estimator of* $\sum_i x_i \tau_i$ *is equal to*

$$(x^\top L^- x) k \sigma^2.$$

In particular, the variance of the best linear unbiased estimator of the simple difference $\tau_i - \tau_j$ *is*

$$V_{ij} = \left(L^-_{ii} + L^-_{jj} - 2L^-_{ij} \right) k \sigma^2.$$

How do we calculate variance?

Theorem

Assume that all the noise is independent, with variance σ^2.
If $\sum_i x_i = 0$, *then the variance of the best linear unbiased estimator of* $\sum_i x_i \tau_i$ *is equal to*

$$(x^\top L^- x) k \sigma^2.$$

In particular, the variance of the best linear unbiased estimator of the simple difference $\tau_i - \tau_j$ *is*

$$V_{ij} = \left(L^-_{ii} + L^-_{jj} - 2L^-_{ij} \right) k \sigma^2.$$

Put $\bar{V} =$ average value of the V_{ij}. Then

$$\bar{V} = \frac{2k\sigma^2 \operatorname{Tr}(L^-)}{v-1} = 2k\sigma^2 \times \frac{1}{\text{harmonic mean of } \theta_1, \ldots, \theta_{v-1}},$$

where $\theta_1, \ldots, \theta_{v-1}$ are the nontrivial eigenvalues of L.

Optimality

The design is called

- A-optimal if it minimizes the average of the variances V_{ij};

over all block designs with block size k and the given v and b.

Optimality

The design is called

- A-optimal if it minimizes the average of the variances V_{ij};

- D-optimal if it minimizes the volume of the confidence ellipsoid for $(\tau_1, \ldots, \tau_v)$;

over all block designs with block size k and the given v and b.

Optimality

The design is called

- A-optimal if it minimizes the average of the variances V_{ij};
  —equivalently, it maximizes the harmonic mean of the non-trivial eigenvalues of the Laplacian matrix L;
- D-optimal if it minimizes the volume of the confidence ellipsoid for $(\tau_1, \ldots, \tau_v)$;

over all block designs with block size k and the given v and b.

Optimality

The design is called

- A-optimal if it minimizes the average of the variances V_{ij};
  —equivalently, it maximizes the harmonic mean of the non-trivial eigenvalues of the Laplacian matrix L;
- D-optimal if it minimizes the volume of the confidence ellipsoid for $(\tau_1, \ldots, \tau_v)$;
  —equivalently, it maximizes the geometric mean of the non-trivial eigenvalues of the Laplacian matrix L;

over all block designs with block size k and the given v and b.

Optimality

The design is called

- A-optimal if it minimizes the average of the variances V_{ij};
  —equivalently, it maximizes the harmonic mean of the non-trivial eigenvalues of the Laplacian matrix L;
- D-optimal if it minimizes the volume of the confidence ellipsoid for $(\tau_1, \ldots, \tau_v)$;
  —equivalently, it maximizes the geometric mean of the non-trivial eigenvalues of the Laplacian matrix L;

over all block designs with block size k and the given v and b.

(Agricultural statisticians tend to favour A-optimality; industrial statisticians prefer D-optimality.)

Balanced designs are optimal

Theorem (Kshirsagar, 1958; Kiefer, 1975)

If there is a balanced incomplete-block design (BIBD) (2-design) for v treatments in b blocks of size k, then it is A- and D-optimal.

Balanced designs are optimal

Theorem (Kshirsagar, 1958; Kiefer, 1975)

If there is a balanced incomplete-block design (BIBD) (2-design) for v treatments in b blocks of size k, then it is A- and D-optimal.

Hence a general idea that

- designs optimal on either of these criteria should be close to balanced
- designs optimal on either of these criteria are not very bad on the other.

D-optimality: spanning trees

A spanning tree for the graph is a collection of edges of the graph which form a tree (graph with no cycles) and which include every vertex.

D-optimality: spanning trees

A spanning tree for the graph is a collection of edges of the graph which form a tree (graph with no cycles) and which include every vertex.

Cheng (1981), after Gaffke (1978), after Kirchhoff (1847):

product of non-trivial eigenvalues of $L = v \times$ number of spanning trees.

So a design is D-optimal iff its concurrence graph has the maximal number of spanning trees.

D-optimality: spanning trees

A spanning tree for the graph is a collection of edges of the graph which form a tree (graph with no cycles) and which include every vertex.

Cheng (1981), after Gaffke (1978), after Kirchhoff (1847):

product of non-trivial eigenvalues of $L = v \times$ number of spanning trees.

So a design is D-optimal iff its concurrence graph has the maximal number of spanning trees.

This is easy to calculate by hand when the graph is sparse.

Electrical networks

We can consider the concurrence graph as an electrical network with a 1-ohm resistance in each edge. Connect a 1-volt battery between vertices i and j. Current flows in the network, according to these rules.

1. Ohm's Law:
   In every edge, voltage drop = current × resistance = current.
2. Kirchhoff's Voltage Law:
   The total voltage drop from one vertex to any other vertex is the same no matter which path we take from one to the other.
3. Kirchhoff's Current Law:
   At every vertex which is not connected to the battery, the total current coming in is equal to the total current going out.

Find the total current I from i to j, then use Ohm's Law to define the effective resistance R_{ij} between i and j as $1/I$.

Electrical networks: A-optimality

Theorem

The effective resistance R_{ij} between vertices i and j is

$$R_{ij} = \left(L^-_{ii} + L^-_{jj} - 2L^-_{ij}\right).$$

Electrical networks: A-optimality

Theorem

The effective resistance R_{ij} between vertices i and j is

$$R_{ij} = \left(L^-_{ii} + L^-_{jj} - 2L^-_{ij}\right).$$

So

$$V_{ij} = R_{ij} \times k\sigma^2.$$

Electrical networks: A-optimality

Theorem

The effective resistance R_{ij} between vertices i and j is

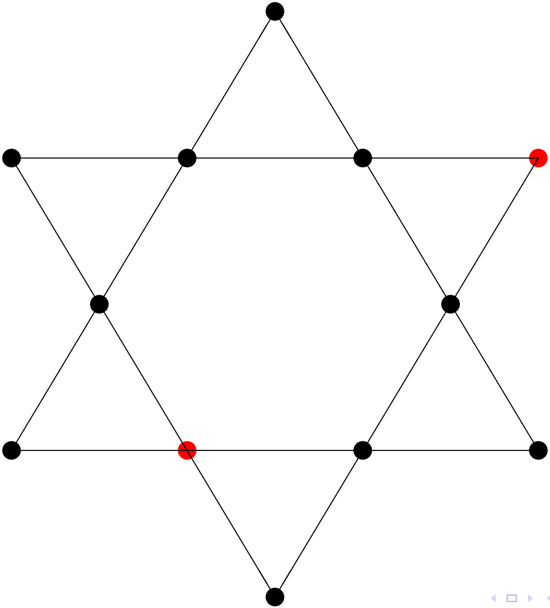
$$R_{ij} = \left(L^-_{ii} + L^-_{jj} - 2L^-_{ij} \right).$$

So

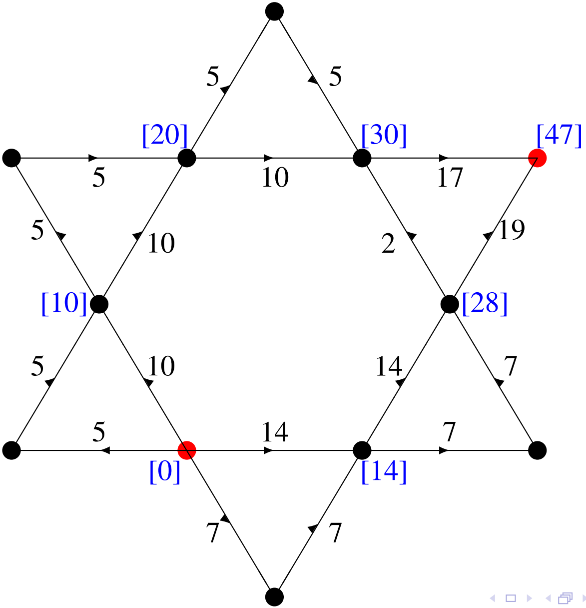
$$V_{ij} = R_{ij} \times k\sigma^2.$$

Effective resistances are easy to calculate without matrix inversion if the graph is sparse.

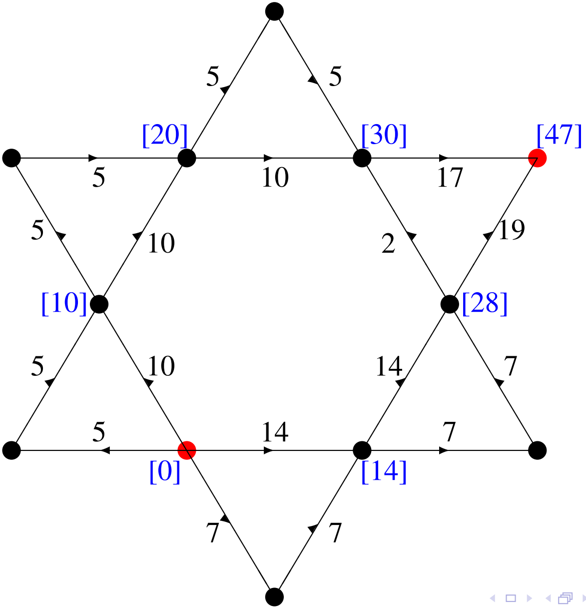
Example calculation



Example calculation



Example calculation



$V = 47$

$I = 36$

$R = \frac{47}{36}$

Some history

- If there is only one block,
  then it is best for the replications to be as equal as possible.

Some history

- If there is only one block,
  then it is best for the replications to be as equal as possible.
- From the early 20-th century, most agricultural experiments had all replications equal to 3 or 4, and $k \geq 3$.

Some history

- If there is only one block,
  then it is best for the replications to be as equal as possible.
- From the early 20-th century, most agricultural experiments had all replications equal to 3 or 4, and $k \geq 3$.
- In the 1930s, 40s and 50s, analysis of data from an experiment involved inverting the Laplacian matrix L without a computer.

Some history

- If there is only one block,
  then it is best for the replications to be as equal as possible.
- From the early 20-th century, most agricultural experiments had all replications equal to 3 or 4, and $k \geq 3$.
- In the 1930s, 40s and 50s, analysis of data from an experiment involved inverting the Laplacian matrix L without a computer.
- In 1958, Kshirsagar published the result that BIBDs are A-optimal among equi-replicate designs; in 1975, Kiefer published the result that they are A- and D-optimal
  (so people tried to make designs as balanced as possible).

Some more history

- In 1982, John and Williams published a short paper in JRSSB on conjectures for optimal block designs, including
  - the set of designs with almost-equal replication will always contain one that is optimal without this restriction
  - the same designs are optimal on the A- and D-criteria.

Some more history

- In 1980, Jones and Eccleston published a short paper in JRSSB on the results of a computer search for A-optimal designs with $k = 2$ and $v = b \leq 10$ (so average replication $= \bar{r} = 2$); when $v = 9$ and $v = 10$ the optimal design is almost a queen-bee design.
- In 1982, John and Williams published a short paper in JRSSB on conjectures for optimal block designs, including
  - the set of designs with almost-equal replication will always contain one that is optimal without this restriction
  - the same designs are optimal on the A- and D-criteria.

Some more recent history

- At the end of the 20-th century, there was an explosion of experiments in genomics, using microarrays. These are effectively block designs with $k = 2$, and biologists wanted A-optimal designs, but they did not know the vocabulary "block" or "A-optimal".

Some more recent history

- At the end of the 20-th century, there was an explosion of experiments in genomics, using microarrays. These are effectively block designs with $k = 2$, and biologists wanted A-optimal designs, but they did not know the vocabulary "block" or "A-optimal".
- In 2001, Kerr and Churchill published the results of a computer search for A-optimal designs with $k = 2$ and $v = b \leq 11$.

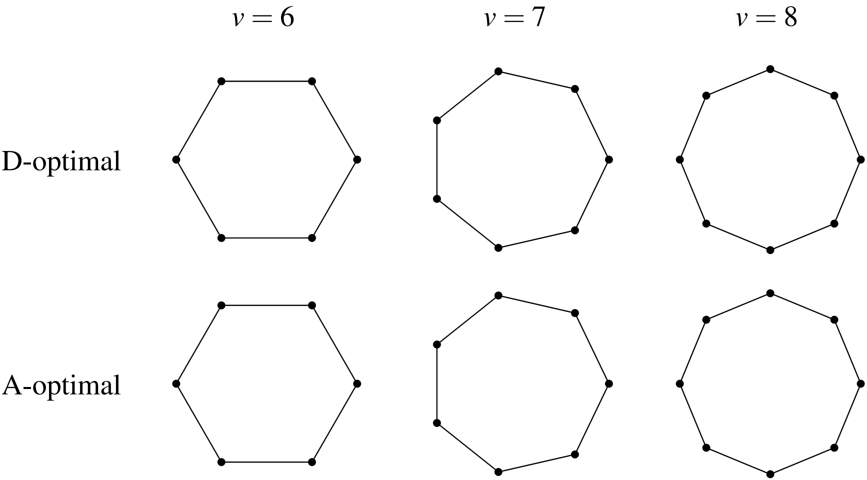
Some more recent history

- At the end of the 20-th century, there was an explosion of experiments in genomics, using microarrays. These are effectively block designs with $k = 2$, and biologists wanted A-optimal designs, but they did not know the vocabulary "block" or "A-optimal".
- In 2001, Kerr and Churchill published the results of a computer search for A-optimal designs with $k = 2$ and $v = b \leq 11$.
- In 2005, Wit, Nobile and Khanin published a paper in JRSSC giving the results of a computer search for A- and D-optimal designs with $k = 2$ and $v = b \leq 13$. The A-optimal designs differ from the D-optimal designs when $v \geq 9$.

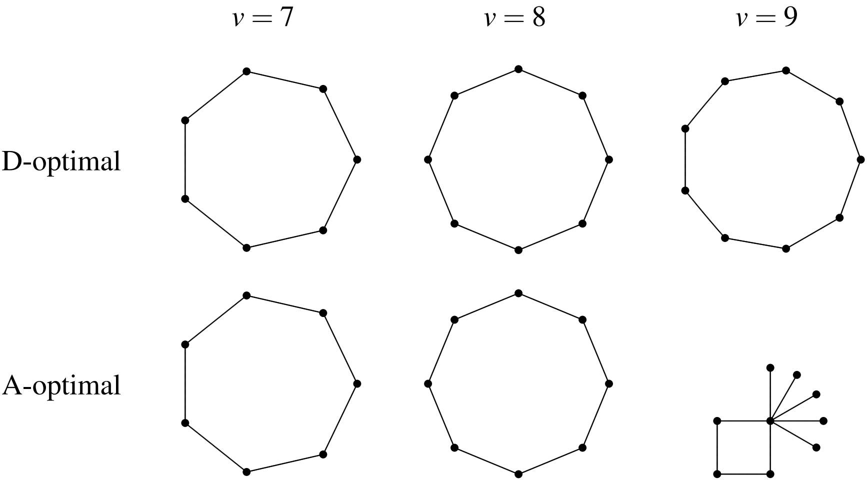
Some more recent history

- At the end of the 20-th century, there was an explosion of experiments in genomics, using microarrays. These are effectively block designs with $k = 2$, and biologists wanted A-optimal designs, but they did not know the vocabulary "block" or "A-optimal".
- In 2001, Kerr and Churchill published the results of a computer search for A-optimal designs with $k = 2$ and $v = b \leq 11$.
- In 2005, Wit, Nobile and Khanin published a paper in JRSSC giving the results of a computer search for A- and D-optimal designs with $k = 2$ and $v = b \leq 13$. The A-optimal designs differ from the D-optimal designs when $v \geq 9$.
- RAB got annoyed at the lack of proof, and published a paper in JRSSC in 2007 giving the A- and D-optimal designs with $k = 2$ and $v = b$ or $v = b + 1$, for all v.

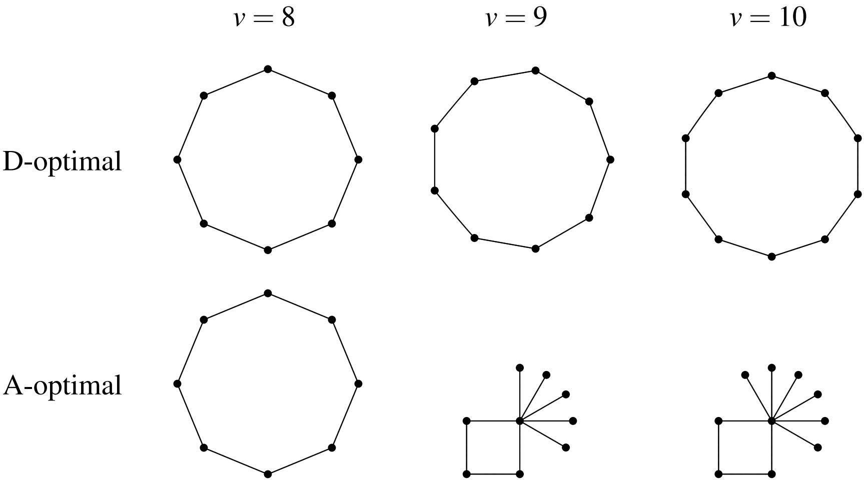
Optimal designs when $k = 2$ and $b = v$



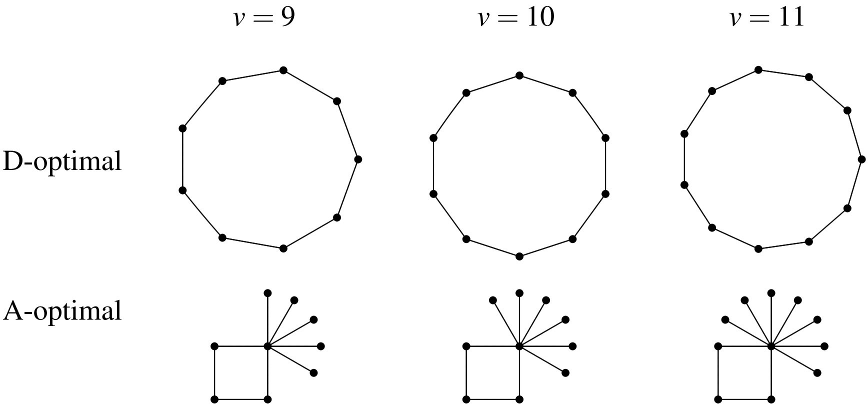
Optimal designs when $k = 2$ and $b = v$



Optimal designs when $k = 2$ and $b = v$



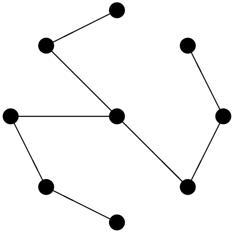
Optimal designs when $k = 2$ and $b = v$

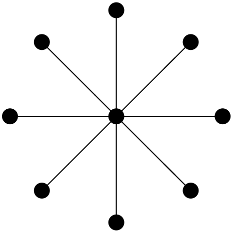


Block size 2: least replication

If $k = 2$ then the design is the same as its concurrence graph, and connectivity requires $b \geq v - 1$.

If $b = v - 1$ then all connected designs are trees.

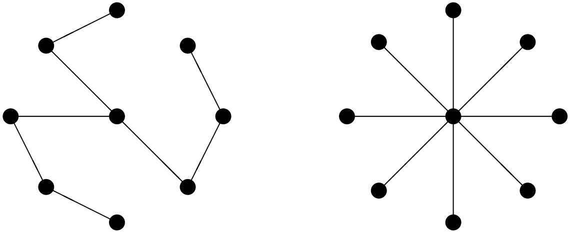




Block size 2: least replication

If $k = 2$ then the design is the same as its concurrence graph, and connectivity requires $b \geq v - 1$.

If $b = v - 1$ then all connected designs are trees.

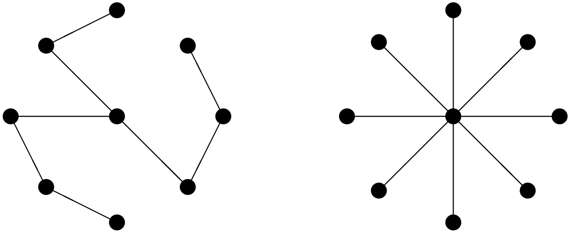


The D-criterion does not differentiate them.

Block size 2: least replication

If $k = 2$ then the design is the same as its concurrence graph, and connectivity requires $b \geq v - 1$.

If $b = v - 1$ then all connected designs are trees.

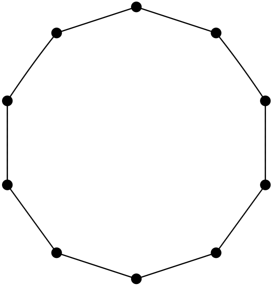


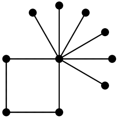
The D-criterion does not differentiate them.

The only A-optimal designs are the stars.

Block size 2: one more block: D

If $k = 2$ and $b = v$ then the design consists of a cycle with trees attached to some vertices.

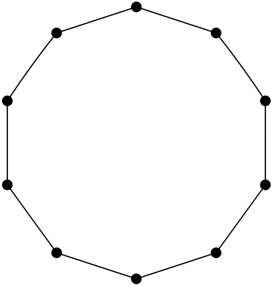


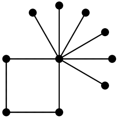


Block size 2: one more block: D

If $k = 2$ and $b = v$ then the design consists of a cycle with trees attached to some vertices.

For a spanning tree, remove one edge without disconnecting the graph.

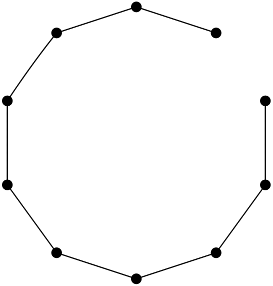


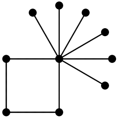


Block size 2: one more block: D

If $k = 2$ and $b = v$ then the design consists of a cycle with trees attached to some vertices.

For a spanning tree, remove one edge without disconnecting the graph.

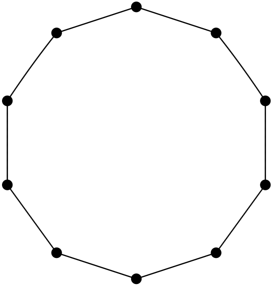


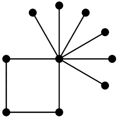


Block size 2: one more block: D

If $k = 2$ and $b = v$ then the design consists of a cycle with trees attached to some vertices.

For a spanning tree, remove one edge without disconnecting the graph.

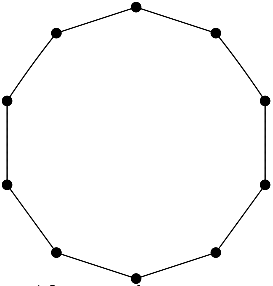




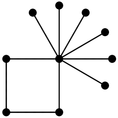
Block size 2: one more block: D

If $k = 2$ and $b = v$ then the design consists of a cycle with trees attached to some vertices.

For a spanning tree, remove one edge without disconnecting the graph.



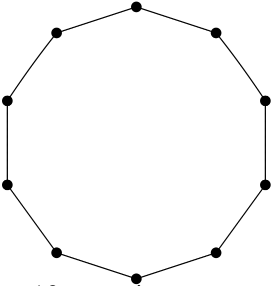
10 spanning trees



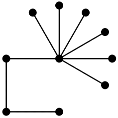
Block size 2: one more block: D

If $k = 2$ and $b = v$ then the design consists of a cycle with trees attached to some vertices.

For a spanning tree, remove one edge without disconnecting the graph.



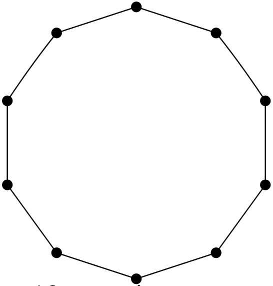
10 spanning trees



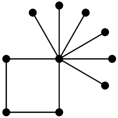
Block size 2: one more block: D

If $k = 2$ and $b = v$ then the design consists of a cycle with trees attached to some vertices.

For a spanning tree, remove one edge without disconnecting the graph.



10 spanning trees

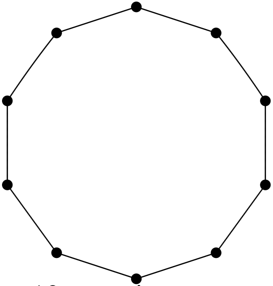


4 spanning trees

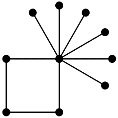
Block size 2: one more block: D

If $k = 2$ and $b = v$ then the design consists of a cycle with trees attached to some vertices.

For a spanning tree, remove one edge without disconnecting the graph.



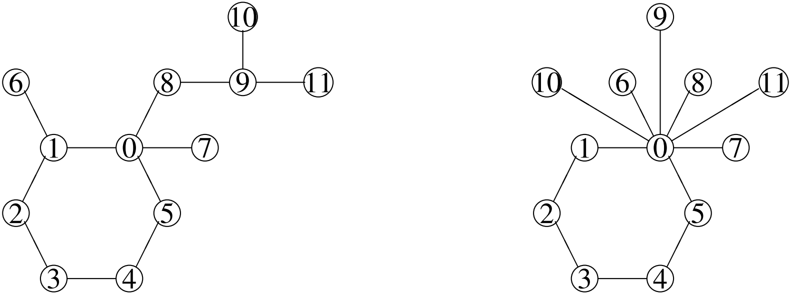
10 spanning trees



4 spanning trees

The cycle is uniquely D-optimal when $b = v$.

Block size 2: one more block: A



For a given size of cycle,
the total variance is minimized
when everything outside the cycle is attached as a leaf to the same vertex of the cycle.

Block size 2: one more block: A

Consider a cycle of length s
with $v - s$ leaves attached to one vertex of the cycle.

Block size 2: one more block: A

Consider a cycle of length s
with $v - s$ leaves attached to one vertex of the cycle.

The sum of the pairwise effective resistances is a cubic function of s with a local minimum in $[2, 5]$ and decreasing with large s.

Block size 2: one more block: A

Consider a cycle of length s
with $v - s$ leaves attached to one vertex of the cycle.

The sum of the pairwise effective resistances is a cubic function of s with a local minimum in $[2, 5]$ and decreasing with large s.

When v is small, the minimum on $[2, v]$ is at v;
for larger v, the minimum on $[2, v]$ is the local minimum.

Block size 2: one more block

D-optimal designs	cycle	always
A-optimal designs	cycle	if $v \leq 8$
	square with leaves attached	if $9 \leq v \leq 12$
	triangle with leaves attached	if $12 \leq v$

Block size 2: one more block

D-optimal designs	cycle	always
A-optimal designs	cycle	if $v \leq 8$
	square with leaves attached	if $9 \leq v \leq 12$
	triangle with leaves attached	if $12 \leq v$

For $v \geq 9$, the ranking on the D-criterion is essentially the opposite of the ranking on the A-criterion,

Block size 2: one more block

D-optimal designs	cycle	always
A-optimal designs	cycle	if $v \leq 8$
	square with leaves attached	if $9 \leq v \leq 12$
	triangle with leaves attached	if $12 \leq v$

For $v \geq 9$, the ranking on the D-criterion is essentially the opposite of the ranking on the A-criterion,
and the A-optimal designs are far from equi-replicate.

Block size 2: one more block

D-optimal designs	cycle	always
A-optimal designs	cycle	if $v \leq 8$
	square with leaves attached	if $9 \leq v \leq 12$
	triangle with leaves attached	if $12 \leq v$

For $v \geq 9$, the ranking on the D-criterion is essentially the opposite of the ranking on the A-criterion,
and the A-optimal designs are far from equi-replicate.
The change is sudden, not gradual.

History again

The A-optimal designs for $k = 2$ and $v = b + 1$ had been given by

- Bapat and Dey in JSPI in 1991
- Mandal, Shah and Sinha in *Calcutta Statistical Assoc. Bull.* in 1991.

History again

The A-optimal designs for $k = 2$ and $v = b + 1$ had been given by

- Bapat and Dey in JSPI in 1991
- Mandal, Shah and Sinha in *Calcutta Statistical Assoc. Bull.* in 1991.

The A-optimal designs for $k = 2$ and $v = b$ had been given by Tjur in *Annals of Statistics* in 1991.

A statistician says . . .

An old collaborator, 1980s

"We all know that the A-optimal designs are essentially the same as the D-optimal designs.
Surely you've got enough mathematics to prove this?"

A statistician says ...

An old collaborator, 1980s

"We all know that the A-optimal designs are essentially the same as the D-optimal designs.
Surely you've got enough mathematics to prove this?"

That old collaborator, December 2008

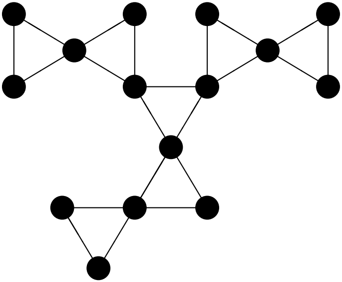
"It seems to be just block size 2 *that is a problem."*

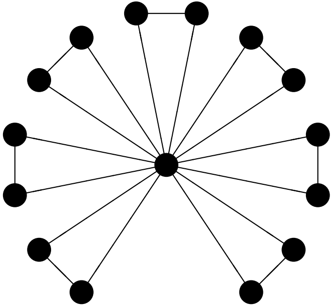
Block size 3, but minimal b

The remaining arguments extend easily to general block size.

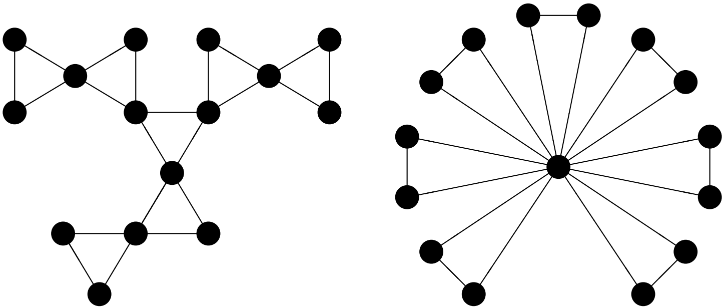
When $k = 3$, for a connected design, we need $2b \geq v - 1$.

If $2b + 1 = v$ then all designs are gum-trees, in the sense that there is a unique sequence of blocks from any one treatment to another.

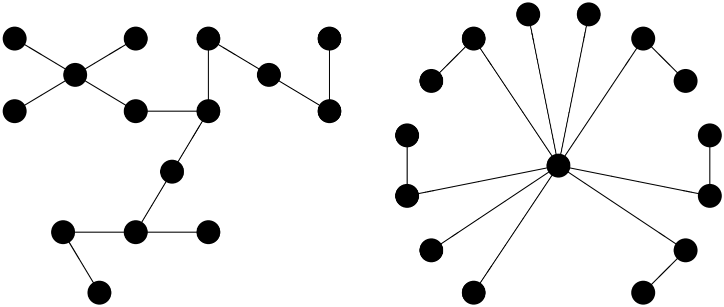




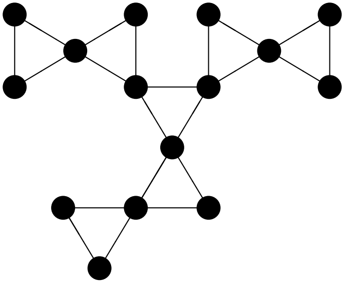
Block size 3, but minimal b: D



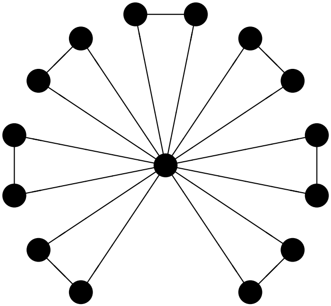
Block size 3, but minimal b: D



Block size 3, but minimal b: D

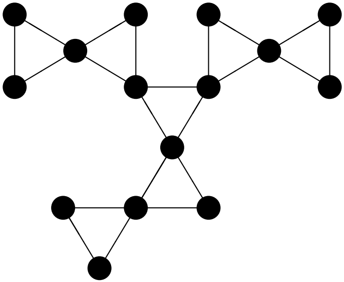


3^7 spanning trees

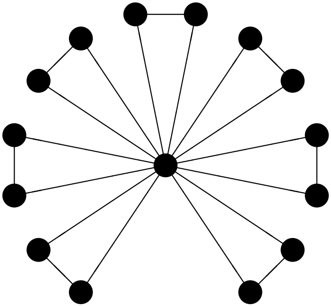


3^7 spanning trees

Block size 3, but minimal b: D



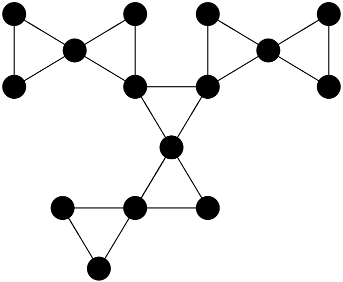
3^7 spanning trees

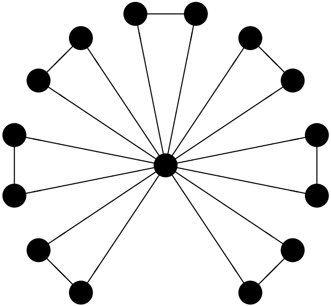


3^7 spanning trees

Every gum-tree with b blocks of size 3 has 3^b spanning trees.
The D-criterion does not differentiate them.

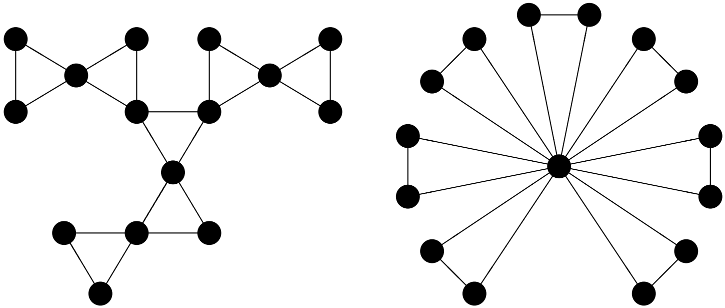
Block size 3, but minimal b: A





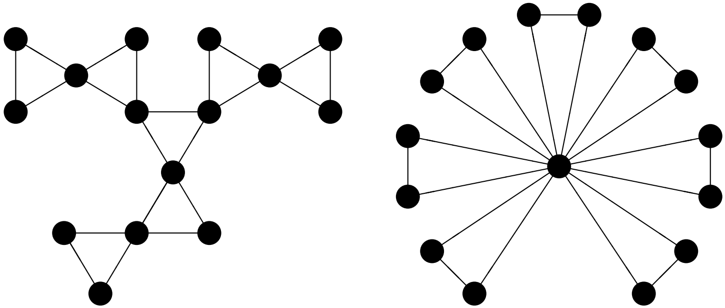
Let R_{ij} be the effective resistance between treatments i and j.

Block size 3, but minimal b: A



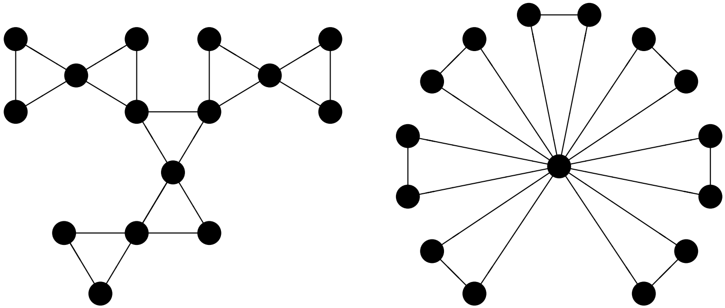
Let R_{ij} be the effective resistance between treatments i and j.
If $i, j \in$ same block then $R_{ij} = \frac{2}{3}$.

Block size 3, but minimal b: A



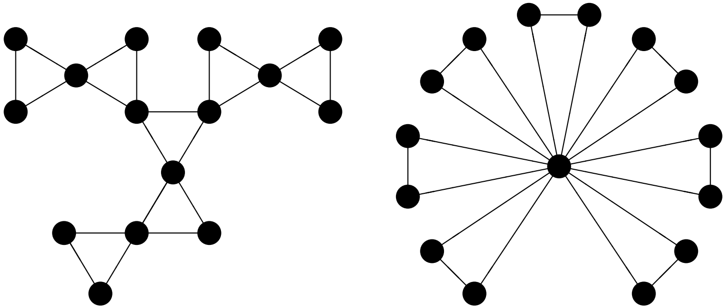
Let R_{ij} be the effective resistance between treatments i and j.
If $i, j \in$ same block then $R_{ij} = \frac{2}{3}$.
If $i, j \in$ distinct intersecting blocks then $R_{ij} = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}$.

Block size 3, but minimal b: A



Let R_{ij} be the effective resistance between treatments i and j.
If $i, j \in$ same block then $R_{ij} = \frac{2}{3}$.
If $i, j \in$ distinct intersecting blocks then $R_{ij} = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}$.
Otherwise, $R_{ij} \geq \frac{2}{3} + \frac{2}{3} + \frac{2}{3} = \frac{6}{3}$.

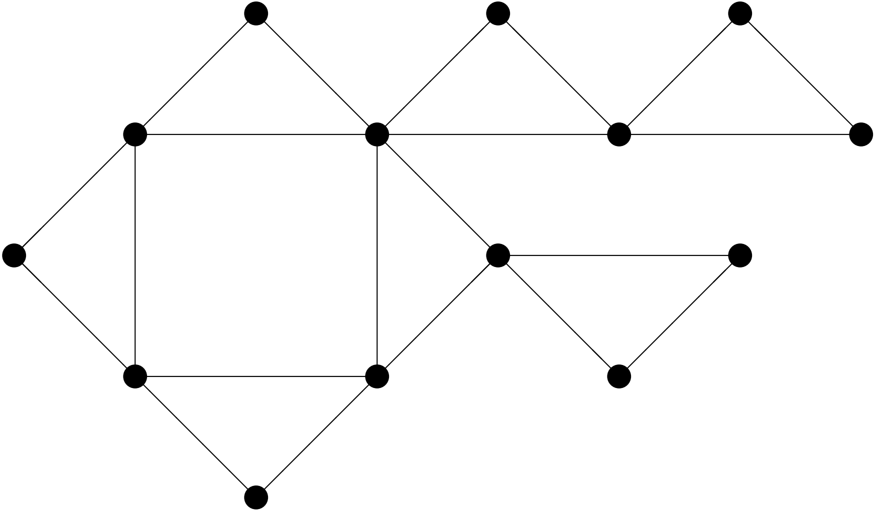
Block size 3, but minimal b: A



Let R_{ij} be the effective resistance between treatments i and j.
If $i, j \in$ same block then $R_{ij} = \frac{2}{3}$.
If $i, j \in$ distinct intersecting blocks then $R_{ij} = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}$.
Otherwise, $R_{ij} \geq \frac{2}{3} + \frac{2}{3} + \frac{2}{3} = \frac{6}{3}$.
The only A-optimal designs are the queen-bee designs.

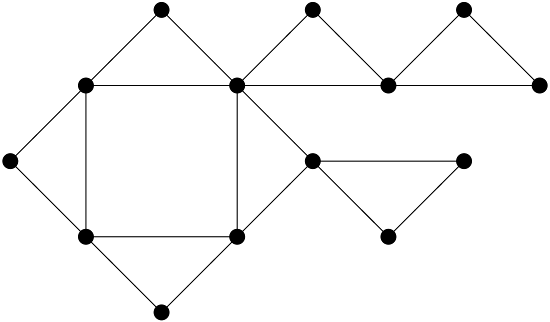
Block size 3, but $b =$ minimal $+1$

If $2b = v$ then G is a gum-cycle with gum-trees attached.



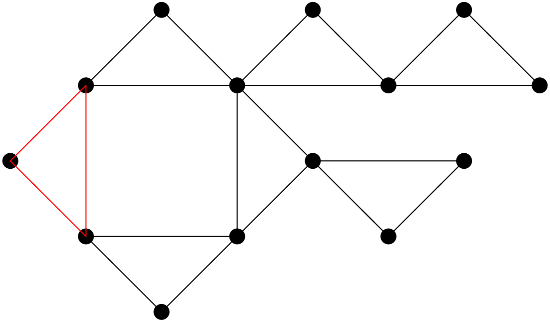
Block size 3, but $b =$ minimal $+1$: D

Suppose that there are s blocks in the gum-cycle.



Block size 3, but $b =$ minimal $+1$: D

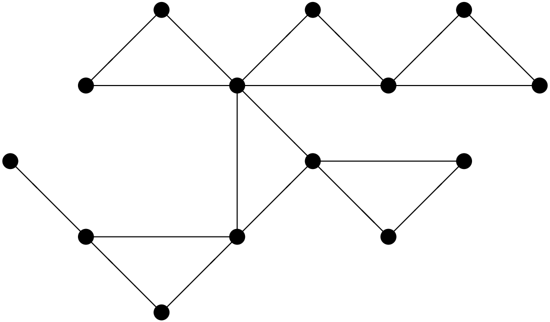
Suppose that there are s blocks in the gum-cycle.



For a spanning tree:
choose a block in the gum-cycle $\qquad s$

Block size 3, but $b =$ minimal $+1$: D

Suppose that there are s blocks in the gum-cycle.

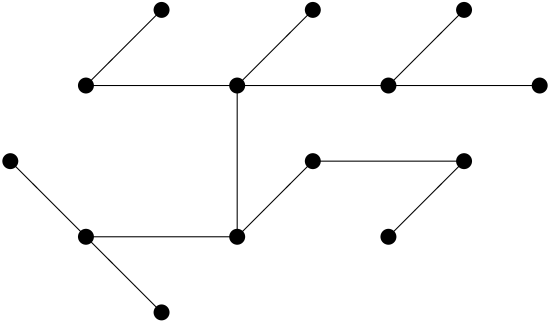


For a spanning tree:

choose a block in the gum-cycle	s
remove its central edge and one other	2

Block size 3, but $b =$ minimal $+1$: D

Suppose that there are s blocks in the gum-cycle.

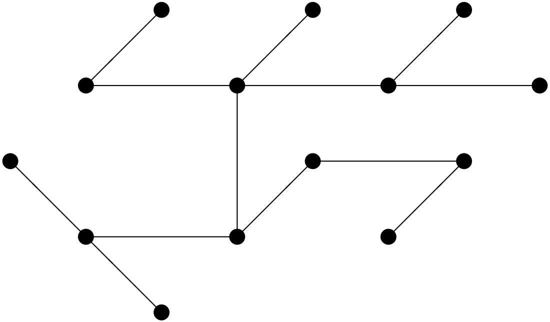


For a spanning tree:

choose a block in the gum-cycle	s
remove its central edge and one other	2
remove an edge from each other block	3^{b-1}

Block size 3, but $b =$ minimal $+1$: D

Suppose that there are s blocks in the gum-cycle.



For a spanning tree:

choose a block in the gum-cycle	s
remove its central edge and one other	2
remove an edge from each other block	3^{b-1}

There are $2s \times 3^{b-1}$ spanning trees.

This is maximized when $s = b$.

Block size k, but $b =$ minimal $+1$: D-optimality

This argument extends to all block sizes.

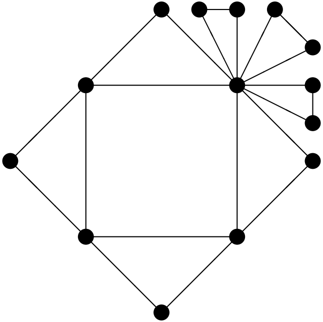
If $v = b(k-1)$ then the only D-optimal designs are the gum-cycles.

Block size 3, but $b =$ minimal $+1$: A

Suppose that there are s blocks in the gum-cycle.

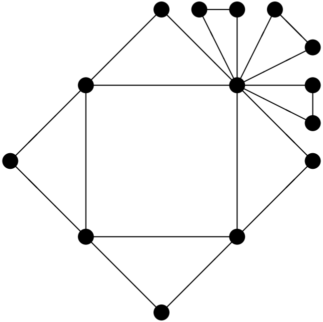
Block size 3, but $b =$ minimal $+1$: A

Suppose that there are s blocks in the gum-cycle.
Then the only candidate for A-optimality consists of $b - s$ triangles attached to a central vertex of the gum-cycle.



Block size 3, but $b =$ minimal $+1$: A

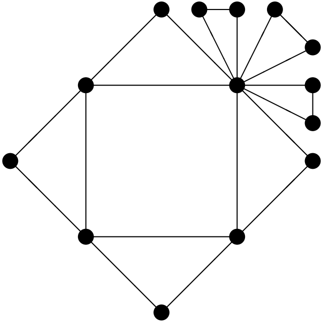
Suppose that there are s blocks in the gum-cycle.
Then the only candidate for A-optimality consists of $b-s$ triangles attached to a central vertex of the gum-cycle.



The sum of the pairwise effective resistances is a cubic function of s.

Block size 3, but $b =$ minimal $+1$: A

Suppose that there are s blocks in the gum-cycle.
Then the only candidate for A-optimality consists of $b - s$ triangles attached to a central vertex of the gum-cycle.

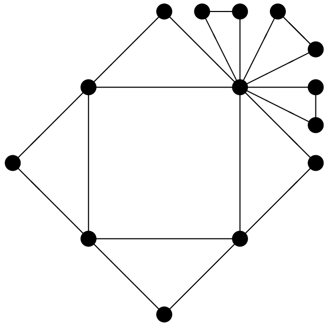


The sum of the pairwise effective resistances is a cubic function of s.

The location of the minimum on $[2, b]$ depends on the value of b.

Block size 3, but $b =$ minimal $+1$: A

Suppose that there are s blocks in the gum-cycle.
Then the only candidate for A-optimality consists of $b - s$ triangles attached to a central vertex of the gum-cycle.



The sum of the pairwise effective resistances is a cubic function of s.

The location of the minimum on $[2, b]$ depends on the value of b.

A-optimal designs do not have s large.

Block size k, but $b =$ minimal $+1$: A-optimality

This argument extends to all block sizes.

If $v = b(k-1)$ then the only A-optimal designs consist of
a gum-cycle of s_0 blocks together with
$b - s_0$ blocks attached to a central vertex of the gum-cycle.

The value of s_0 depends on b and k, but it is never large.

Block size k, but $b =$ minimal $+1$: A-optimality

This argument extends to all block sizes.

If $v = b(k-1)$ then the only A-optimal designs consist of
a gum-cycle of s_0 blocks together with
$b - s_0$ blocks attached to a central vertex of the gum-cycle.

The value of s_0 depends on b and k, but it is never large.

k \ b	2	3	4	5	6	7	8	9	10	11	12	13
2	2	3	4	5	6	7	8	4	4	4	3 or 4	3
3	2	3	4	5	6	3	3	3	3	3	2	2
4	2	3	4	5	3	2	2	2	2	2	2	2
5	2	3	4	5	2	2	2	2	2	2	2	2
6	2	3	4	2	2	2	2	2	2	2	2	2

Block size k, but $b =$ minimal $+1$: A-optimality

This argument extends to all block sizes.

If $v = b(k-1)$ then the only A-optimal designs consist of
a gum-cycle of s_0 blocks together with
$b - s_0$ blocks attached to a central vertex of the gum-cycle.

The value of s_0 depends on b and k, but it is never large.

k b	2	3	4	5	6	7	8	9	10	11	12	13
2	2	3	4	5	6	7	8	4	4	4	3 or 4	3
3	2	3	4	5	6	3	3	3	3	3	2	2
4	2	3	4	5	3	2	2	2	2	2	2	2
5	2	3	4	5	2	2	2	2	2	2	2	2
6	2	3	4	2	2	2	2	2	2	2	2	2

My PhD student Alia Sajjad discovered that this had been published
by Krafft and Schaefer in JSPI in 1997

Block size k, but $b =$ minimal $+1$: A-optimality

This argument extends to all block sizes.

If $v = b(k-1)$ then the only A-optimal designs consist of
a gum-cycle of s_0 blocks together with
$b - s_0$ blocks attached to a central vertex of the gum-cycle.

The value of s_0 depends on b and k, but it is never large.

k \ b	2	3	4	5	6	7	8	9	10	11	12	13
2	2	3	4	5	6	7	8	4	4	4	3 or 4	3
3	2	3	4	5	6	3	3	3	3	3	2	2
4	2	3	4	5	3	2	2	2	2	2	2	2
5	2	3	4	5	2	2	2	2	2	2	2	2
6	2	3	4	2	2	2	2	2	2	2	2	2

My PhD student Alia Sajjad discovered that this had been published
by Krafft and Schaefer in JSPI in 1997
(but they did not know about Tjur, 1991).

New asymptotic results (large v)

Current work by J. Robert Johnson and Mark Walters.

Block size 2; one control treatment; want to minimize the average variance of comparisons with control.

New asymptotic results (large v)

Current work by J. Robert Johnson and Mark Walters.

Block size 2; one control treatment; want to minimize the average variance of comparisons with control.

Average replication	optimal design (probably)
2 and a little above	many small designs (including many leaves) glued at the control
around 3	one large random almost-regular graph with average replication 3.5, also quite a lot of edges from points in this to the control, and a bunch of leaves rooted at the control
4 and above	a random almost-regular graph (maybe with a few leaves)

Agricultural trials: I lied

In breeding trials of new varieties, typically there is very little seed of each of the new varieties.

Tradtionally, an experiment has one plot for each new variety and several plots for a well-established but uninteresting “control”: for example, 30 new varieties on one plot each and one control on 6 plots.

Agricultural trials: I lied

In breeding trials of new varieties, typically there is very little seed of each of the new varieties.

Tradtionally, an experiment has one plot for each new variety and several plots for a well-established but uninteresting "control": for example, 30 new varieties on one plot each and one control on 6 plots.

In the last 10 years, Cullis and colleagues have suggested replacing the control by double replicates of a small number of new varieties: for example, 24 new varieties with one plot each and 6 new varieties with two plots each.

Agricultural trials: I lied

In breeding trials of new varieties, typically there is very little seed of each of the new varieties.

Tradtionally, an experiment has one plot for each new variety and several plots for a well-established but uninteresting “control”: for example, 30 new varieties on one plot each and one control on 6 plots.

In the last 10 years, Cullis and colleagues have suggested replacing the control by double replicates of a small number of new varieties: for example, 24 new varieties with one plot each and 6 new varieties with two plots each.

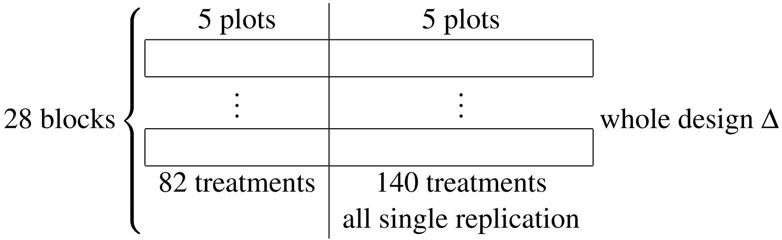
This is an improvement if there are no blocks.

Current work

Brian Cullis and David Butler:
the milling phase of a wheat variety trial has 222 treatments in 28 blocks of size 10. ($280 - 222 = 58$ and $222 - 58 = 164$, so at least 164 treatments must have single replication.)

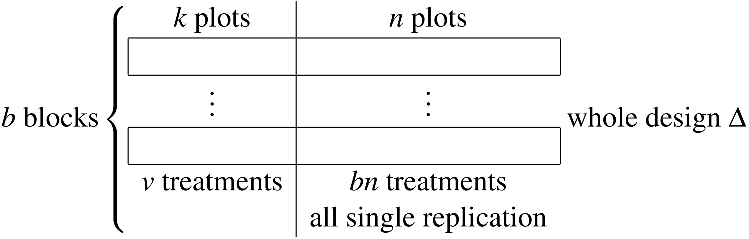
Current work

Brian Cullis and David Butler:
the milling phase of a wheat variety trial has 222 treatments in 28 blocks of size 10. ($280 - 222 = 58$ and $222 - 58 = 164$, so at least 164 treatments must have single replication.)



subdesign Γ has 82 treatments
in 28 blocks of size 5

A general block design with average replication less than 2



Whole design Δ has $v+bn$ treatments in b blocks of size $k+n$;
the subdesign Γ has v treatments in b blocks of size k.

Current work

Theorem (cf Martin, Chauhan, Eccleston and Chan, 2006; Herzberg and Jarrett, 2007)

The sum of the variances of treatment differences in Δ

$$= constant + V_1 + nV_3 + n^2 V_2,$$

where

$$\begin{aligned} V_1 &= \text{the sum of the variances of treatment differences in } \Gamma \\ V_2 &= \text{the sum of the variances of block differences in } \Gamma \\ V_3 &= \text{the sum of the variances of sums of} \\ &\quad\ \text{one treatment and one block in } \Gamma. \end{aligned}$$

(If Γ is equi-replicate then V_2 and V_3 are increasing functions of V_1.)

Theorem (cf Martin, Chauhan, Eccleston and Chan, 2006; Herzberg and Jarrett, 2007)

The sum of the variances of treatment differences in Δ

$$= constant + V_1 + nV_3 + n^2 V_2,$$

where

$$\begin{aligned} V_1 &= \textit{the sum of the variances of treatment differences in } \Gamma \\ V_2 &= \textit{the sum of the variances of block differences in } \Gamma \\ V_3 &= \textit{the sum of the variances of sums of} \\ &\quad\ \textit{one treatment and one block in } \Gamma. \end{aligned}$$

(If Γ is equi-replicate then V_2 and V_3 are increasing functions of V_1.)

Consequence

For a given choice of k, make Γ *as efficient as possible.*

Current work

Consequence

If n or b is large,
it may be best to make Γ a complete block design for k' controls, even if there is no interest in comparisons between new treatments and controls, or between controls.

$5n+10$ treatments in 5 blocks of size $4+n$

1	2	3	4	A_1	$\cdots$	A_n
3	4	5	6	B_1	$\cdots$	B_n
5	6	7	8	C_1	$\cdots$	C_n
7	8	9	0	D_1	$\cdots$	D_n
9	0	1	2	E_1	$\cdots$	E_n

Youden and Connor (1953): "experiments in physics do not need much replication because results are not very variable" — chain block design

$5n+10$ treatments in 5 blocks of size $4+n$

1	2	3	4	A_1	$\cdots$	A_n
3	4	5	6	B_1	$\cdots$	B_n
5	6	7	8	C_1	$\cdots$	C_n
7	8	9	0	D_1	$\cdots$	D_n
9	0	1	2	E_1	$\cdots$	E_n

Youden and Connor (1953): "experiments in physics do not need much replication because results are not very variable" — chain block design

1	2	3	4	A_1	$\cdots$	A_n
1	5	6	7	B_1	$\cdots$	B_n
2	5	8	9	C_1	$\cdots$	C_n
3	6	8	0	D_1	$\cdots$	D_n
4	7	9	0	E_1	$\cdots$	E_n

subdesign is dual of BIBD (Herzberg and Andrews, 1978)

$5n+10$ treatments in 5 blocks of size $4+n$

1	2	3	4	A_1	$\cdots$	A_n
1	5	6	7	B_1	$\cdots$	B_n
2	5	8	9	C_1	$\cdots$	C_n
3	6	8	0	D_1	$\cdots$	D_n
4	7	9	0	E_1	$\cdots$	E_n

subdesign is dual of BIBD

$5n+10$ treatments in 5 blocks of size $4+n$

1	2	3	4	A_1	$\cdots$	A_n
1	5	6	7	B_1	$\cdots$	B_n
2	5	8	9	C_1	$\cdots$	C_n
3	6	8	0	D_1	$\cdots$	D_n
4	7	9	0	E_1	$\cdots$	E_n

subdesign is dual of BIBD

1	2	3	6	A_1	$\cdots$	A_n
2	3	4	7	B_1	$\cdots$	B_n
3	4	5	8	C_1	$\cdots$	C_n
4	5	1	9	D_1	$\cdots$	D_n
5	1	2	0	E_1	$\cdots$	E_n

best subdesign for $k=3$
is better for large n if $b \neq 5$

$5n+10$ treatments in 5 blocks of size $4+n$

1	2	3	6	A_1	$\cdots$	A_n
2	3	4	7	B_1	$\cdots$	B_n
3	4	5	8	C_1	$\cdots$	C_n
4	5	1	9	D_1	$\cdots$	D_n
5	1	2	0	E_1	$\cdots$	E_n

best subdesign for $k=3$
is better for large n if $b \neq 5$

$5n+10$ treatments in 5 blocks of size $4+n$

1	2	3	6	A_1	$\cdots$	A_n
2	3	4	7	B_1	$\cdots$	B_n
3	4	5	8	C_1	$\cdots$	C_n
4	5	1	9	D_1	$\cdots$	D_n
5	1	2	0	E_1	$\cdots$	E_n

best subdesign for $k=3$
is better for large n if $b \neq 5$

K_1	K_2	1	2	A_1	$\cdots$	A_n
K_1	K_2	3	4	B_1	$\cdots$	B_n
K_1	K_2	5	6	C_1	$\cdots$	C_n
K_1	K_2	7	8	D_1	$\cdots$	D_n
K_1	K_2	9	0	E_1	$\cdots$	E_n

better for large n if $b > 13$ even if there is no interest in controls