Optimal Design of Experiments on Connected Units with Application to Social Networks

Ben M. Parker[*1], Steven G. Gilmour [2] and John Schormans[3]

[1]Southampton Statistical Sciences Research Institute, University of Southampton, UK
[2]Department of Mathematics, King's College London, UK
[3]School of Electronic Engineering and Computer Sciences, Queen Mary University of London, UK

June 22, 2016

Abstract

When experiments are performed on social networks, it is difficult to justify the usual assumption of treatment-unit additivity, due to the connections between actors in the network. We investigate how connections between experimental units affect the design of experiments on those experimental units. Specifically, where we have unstructured treatments, whose effects propagate according to a linear network effects model which we introduce, we show that optimal designs are no longer necessarily balanced; we further demonstrate how experiments which do not take a network effect into account can lead to much higher variance than necessary and/or a large bias. We show the use of this methodology in a very wide range of experiments in agricultural trials, and crossover trials, as well as experiments on connected individuals in a social network.

1 Introduction

We investigate how the structure of relationships between experimental units affects the design of experiments on these experimental units. We consider examples where the treatment applied to a single experimental unit has an effect on further experimental units which have some relationship to the one given the treatment. Previous research has focused on agricultural trials, where the treatment applied to an experimental unit (a plot) also has an effect on neighbouring experimental units within a block, and on cross-over experiments where a treatment continues to have an effect on the subject when measured at subsequent time periods. That research generally considers a class of experiments in which the subjects are related according to some regular pattern. Here we remove any restrictions on the regularity of the relationships.

---

[*]B.M.Parker@soton.ac.uk

Our primary application is in investigating experiments on social networks. These networks are prevalent in virtual online communities, such as Facebook or Twitter, but the methodology also encompasses networks in real world communities, such as friendship, familial relationship, or geographical networks. We consider the case where an experiment is to be performed on a known social network and it is expected that the treatments applied will affect not only the nodes of the network to which they are applied, but also all those nodes which are attached to them. This makes it difficult to make the usual assumption in randomized experiments that treatment effects are additive, i.e. the treatment effects have the same distribution no matter which experimental unit they are applied to. This idea is familiar from cross-over trials with carry-over effects and agricultural field experiments with neighbour effects, but the neighbour structures in these applications are usually very regular. Social networks typically have irregular, unbalanced patterns of connections and dealing with this requires some new design methodology. We also adapt this research on social networks to non-standard examples in finding the optimal design of experiments for agricultural trials and crossover trials.

1.1 Motivating examples

Both examples are intended to be illustrative, and are simplified versions of experiments from personal communications of the authors with (1) a marketing consultancy and (2) an academic public health researcher.

1.1.1 Online marketing

In online marketing experiments different advertisements, or different versions of the same advertisement, are investigated to find out which are most effective. For example, an advertisement for a product may describe it as either "half price" or "50% off", or some other alternative phrasings, which we define as treatments in the usual terminology. In other cases the treatments might have a factorial structure, though we will not deal with that case here.

We assume we have a network of 10000 customers, and want to assess the effectiveness of sending 4 different adverts to four subsets each of 25 customers (before we send the best advertisement to the rest of the customers). An advertisement is deemed effective if more customers click on a hyperlink in a message. The advertisements will be released to a test panel of subjects, with each subject being allocated a single treatment. The response of these subjects to the advert will be measured as the number of times that hyperlinks are clicked in the advertisement. (Alternatively we might choose to measure a longer term response, such as a measure of changes in the behaviour of the subjects.)

The standard randomised experiment requires that responses from subjects are independent; in reality, especially with online marketing, advertising to one person will often mean that the advertisement is passed to a second person. Indeed, this viral effect is sometimes exactly what marketers wish to encourage: advertisers pay for distribution to an original population, with popular adverts distributed for free to those in the original target's social network.

1.1.2 Public Health Experiment

In a public health experiment, nutritionists were interested in the effect of different interventions for promoting weight loss amongst a group of volunteers that knew each other socially, according to some friendship network structure. For example, some subjects might be sent daily information by text message about healthy eating, and some might be sent a weekly magazine with low fat recipes. As well as the effectiveness of the interventions, researchers were interested in whether the message sent to one subject had an effect on other subjects within the trial who were connected within the original subject's social network.

1.2 Outline of this paper

We set out in this paper to investigate how experimental design might take into account the structure of a social network, to obtain the optimal information about which adverts are best when experimental units are not independent. We see that these social networks readily generalise to enable us to consider a framework which applies to many situations for designing experiments on connected experimental units.

In Section 2 we set out the model considered in this paper, and review previous work in design for connected experimental units. We introduce terminology and the methodology for designing experiments on this model in Section 3, before showing the effect of misspecification of the model in Section 4. The importance of the network structure is discussed in Section 5, and algorithms for finding an optimal design in Section 6. We present a collection of examples in Section 7, before we conclude briefly in Section 8.

2 Models Considered and Previous Work

We consider networks of subjects; a network $G = (N, E)$ is an undirected graph, a collection of nodes N and edges $E \subseteq (N \times N)$ where the nodes represent experimental units (typically subjects) on which we apply some treatment. The edges represent relationships between the subjects.

We assume that if a relationship exists between two subjects, the response of each subject is dependent on the treatment applied to the other according to a linear network effects model which we specify below.

We have $|N| = n$ subjects, and m treatments. The relationship between subjects is specified by the $n \times n$ adjacency matrix A where $A_{ik} = 1$ if and only if i and k are related and $A_{ik} = 0$ otherwise. By convention, $A_{ii} = 0$. For ease of exposition in this paper, we assume initially that links are non-directional, such that $A_{ik} = A_{ki}$, i.e. A is a symmetric matrix.

We assume there is a "subject effect" on the response from subject i of τ_j when treatment j is given to that subject, and a "network effect" of γ_l if treatment l is given to a connected subject k (if $A_{ik} = 1$). We assume that each subject receives exactly one treatment.

3

2.1 Linear network effects model

We measure the response Y_i for each of our subjects. Let $t(i)$ be the treatment applied to subject i; then our response is modelled as

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

We assume that errors ϵ_i are independent and identically distributed with mean 0 and constant variance σ^2. We assume we wish to estimate the subject and/or network effects, or some contrast of them.

We let $\boldsymbol{u}_j$ be the indicator vector with i-th element equal to 1 when treatment j is applied to subject i, and otherwise 0. (This terminology is derived from that used section 2.7 of Bailey (2008), for example.) In matrix form,

$$\mathbb{E}\,(\boldsymbol{Y}) = \begin{pmatrix} \boldsymbol{1} & \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_m & A\boldsymbol{u}_1 & A\boldsymbol{u}_2 & \dots & A\boldsymbol{u}_m \end{pmatrix} \begin{pmatrix} \mu \\ \boldsymbol{\tau} \\ \boldsymbol{\gamma} \end{pmatrix} = F\boldsymbol{\beta}, \tag{2}$$

where F is our design matrix, and our vector of parameters is

$$\boldsymbol{\beta} = \begin{pmatrix} \mu & \boldsymbol{\tau^T} & \boldsymbol{\gamma^T} \end{pmatrix}^T = (\mu \quad \tau_1 \dots \tau_m \quad \gamma_1 \dots \gamma_m)^T. \tag{3}$$

We let the design space, the set of possible experiments we may perform, be

$$\mathcal{X} = \{[t(1), \dots, t(n)] | t(i) \in \{1, \dots, m\}, i = 1, \dots, n\},$$

such that we give one and only one treatment to each subject. We calculate the Fisher information matrix as $I = F^T F$.

2.2 Previous work in design for connected experimental units

Previous work in experimental design for connected units has been limited, and focused on agricultural trials, where consideration is occasionally given to how the response of one plot is affected by the treatment given to neighbouring plots, as well as the treatment given to the plot itself (which is often called the "direct treatment effect", though we refer to it in this paper as the subject effect).

Besag and Kempton (1986) review the analysis of field experiments where responses in a plot are affected by the treatments applied to neighbouring plots.

Druilhet (1999) considers designing experiments in linear blocks, with block effects and with left and right neighbour effects only in the direction of the blocks, in the context of both agricultural trials, and crossover trials. The model used is $\mathbb{E}\,(\boldsymbol{Y}) = b\boldsymbol{\beta} + T\boldsymbol{\tau} + L\boldsymbol{\lambda} + R\boldsymbol{\rho}$ where $\boldsymbol{\beta}$ are block effects, $\boldsymbol{\tau}$ are treatment effects, $\boldsymbol{\lambda}$ are left-neighbour effects, $\boldsymbol{\rho}$ (sometimes assumed to be zero for uni-directional effects) are right-neighbour effects, and B, T, L, and R, are incidence matrices for block, treatment, and neighbour effects. He discusses circular neighbour-balanced designs, where the number of times treatment i is to the left of treatment

4

j is constant, for $i \neq j$, and shows these are optimal for the estimation of treatment effects under certain conditions.

These two previous papers both assume that there is an edge plot, or guard plot, such that each plot of interest (the interior plots) has two neighbours. Kunert and Martin (2000) relax this assumption, but still assume that there are one-dimensional arrangements of plots in each block (i.e. that a field trial has a left and right neighbour effect, but no up or down effect). They present criteria for designs to be optimal and, for certain examples with block size up to four, demonstrate that there are designs which fulfill these criteria, but do not show how to construct optimal designs. Kunert and Mersmann (2011) extend this to larger block sizes. Azaïs and Druilhet (1997) considers minimising bias in this model, and shows that neighbour-balanced (randomized) designs minimize bias.

The models considered in Druilhet (1999), Kunert and Martin (2000), and Kunert and Mersmann (2011) are, with left and right neighbour effects assumed to be equal and neglecting the block effects, a special case of the Linear Network effects model (equation (1)) we consider in this paper.

Pearce (1957) assumes a treatment affects the plot it is applied to (a "local effect"), and also other plots in the same block (a "remote effect"). He allows interaction between these effects, i.e. the remote effect of a treatment on a plot depends on the treatment applied locally to that plot. He shows orthogonal designs no longer guarantee all effects are estimable, and proposes balanced incomplete block designs with restrictions of the number of treatments to maintain estimability.

In designs for cross-over trials, subjects receive sequences of treatments, and a carry-over effect is often assumed to the next time period. We show in this paper how to represent these cross-over designs using a generalisation of our model; particular examples have been studied. e.g. Hedayat and Zhao (1990) considers two period cross-over designs, and show in particular that optimal designs for estimating the carry-over effects are not always balanced.

These temporal effects can be thought of in the same framework as treatments affecting spatially neighbouring subjects in agricultural experiments. See, for example, Jones and Kenward (2003).

There has been little work discussing designing experiments on subjects that are connected by a social network. In McConnell et al. (2010), the stable unit treatment value assumption (Rubin 1978) is discussed, and the assumption is questioned by detecting whether there is a "spillover effect" in an experiment on voter mobilization where voters are sent mailings before an election. Here a multilevel model is applied and household and neighbourhood effects are detected, where the household and neighbourhood effects have a regular pattern within nested blocks.

2.3 Previous work on inference for social networks

A review of modelling social influence was performed by Leenders (2002); the problem considered is to infer relationships between actors (nodes) in a social network by measuring some response on those networks. This paper considers two different types of models. The first class is of the form

$$\boldsymbol{y} = \rho W \boldsymbol{y} + X \boldsymbol{\beta} + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I). \tag{4}$$

Here ρ is a scalar, and W a matrix representing the strength of relationships between actors.

In this model an actor in a social network forms his response from the (weighted) responses of actors connected to him. This is the network effects model of Doreian et al. (1989), sometimes called a spatial autoregressive model. In a social network, this corresponds to an actor basing his opinion (response) on his own innate response, which is modified by the actual response of his peers. The second class of model considered is

$$\boldsymbol{y} = X\boldsymbol{\beta} + \epsilon, \quad \epsilon = \rho W \epsilon + \nu, \quad \nu \sim N(0, \sigma^2 I). \tag{5}$$

Here the response of an actor in a social network depends on the *change* in response of actors connected to him. This network is sometimes called a spatial moving average model. In social networks, these correspond to an actor altering his opinion (response) due to a change in opinion of his peers; as we demonstrate below, this paradigm seems to be valuable in experiments where changes in responses of an experimental unit may also propagate somewhat to its neighbours because of some spillover or neighbourhood effect caused by applying a treatment.

These two form a wide class of models, and mathematical interest is found in calculating the strength of actors' relationships (the weight matrix W). The non-network effects $\boldsymbol{\beta}$ can be inferred using maximum likelihood and appropriate hypothesis testing carried out to investigate whether a particular network structure exists. The review paper, Leenders (2002), emphasizes the large effect that W has on any conclusions, and the importance of correct model choice.

2.4 Experiments covered in our model and possible generalisations

We focus on the Linear Network effects model (1). Although a relatively simple model, this allows us to demonstrate clearly the optimal design of experiments on networks, a novel methodology. In the methodology described in this paper, and unlike in much previous research in designing experiments, we do not impose any restrictions or symmetries in the relationships between subjects, such that any two subjects may be related or not. We present a linear network effects model for unstructured treatments within this framework, and discuss finding optimal designs for experiments in this case.

This setting admits many experiments, for example in our first motivating example the nodes are subjects in a marketing experiment, and the links are a social relationship, such as being "friends" on Facebook or some other social network. The treatments are different adverts, and the responses the number of hyperlinks clicked by a subject. If a subject is exposed to marketing, their friend may be affected by this marketing and the friend's response may be altered (positively or negatively). If a subject is given some information about a product in a marketing campaign, he may tell a friend about it, which may affect the number of times the friend clicks a hyperlink.

As in our second motivating example, the nodes may be subjects in a public health experiment. The links may be some friendship relationship, some geographical relationship, or some familial relationship. The treatments may be different messages given in the campaign (e.g. "Eat five fruit or vegetables a day") and the response the change in fruits or vegetables eaten after the campaign.

A subject receives a treatment and, whether or not they alter their own behaviour, it may bring about a change in their friend's response.

These examples show that the network effect in the Linear Network effect model is a particular way of representing a social influence, in a similar way to described in section 2.3 above. See section 7 for further examples of the method in agriculture, medicine, and social networks. It is relatively simple to consider generalisations of the Linear Network Effects model. For example, we could normalise the network effect by the number of links, on the basis that the effect of a neighbour on someone with a high number of links is probably smaller than for a sparsely connected individual. We might also impose the restriction that $\gamma_i = \kappa \tau_i$ for some $0 \leq \kappa \leq \tau$ and all i, i.e. that the network effect is a fixed multiple of the subject effect.

We do not consider autoregressive models here. For example, we do not consider experiments where an experimental unit's response is directly affected by its neighbours response, and not affected by the treatment applied to a neighbour. Examples of these include

- a marketing experiment where a subject changes his behaviour based on the response of another customer, and not the treatment applied to a customer. For example, when customers are queueing at the ticket counter of an airport in order to check in their luggage, the service one customer receives is observed by those who are in the queue near him (a directional tie) and is potentially discussed by the by-standers. This affects their perception of the service provided to them by the airline.

- where a child copies his parent's behaviour in response to a public health campaign, this is described by a spatial autoregressive model as the parent's response (and not the parent's treatment) affects the child's response.

Although these generalisations and extensions are, of course, interesting, we seek in this paper to introduce the general technique of designing experiments on social networks with a relatively simple, yet flexible, model, whilst maintaining clarity, and hope that future work will include generalisations to this model. In particular, the generalisation to the spatial autoregressive model adds complexity to the design problem, requiring techniques in non-linear optimal design.

3   Designing Experiments

Writing n_i as the number of subjects given treatment i, n_{ij} as the number of connections between subjects given treatment i and those given treatment j, and $n_{ij}^{(2)}$ as the number of connections between subjects given treatment i and subjects given treatment j with a path of length 2, including trivial paths that use the same edge twice, then we can write the Fisher information matrix as

$$I = F^T F = \begin{pmatrix} n & \mathbf{1}^T X & \mathbf{1}^T A X \\ X^T \mathbf{1} & X^T X & X^T A X \\ X^T A \mathbf{1} & X^T A X & X^T A^2 X \end{pmatrix} \tag{6}$$

where we recall that as A is symmetric, $A^T = A$. This form is useful as it separates the information due to the subject effects and the network effects. Expanding this gives

$$
I = \begin{pmatrix}
n & n_1 & n_2 & \dots & n_m & \sum_i n_{i1} & \sum_i n_{i2} & \dots & \sum_i n_{im} \\
n_1 & n_1 & 0 & \dots & 0 & n_{11} & n_{12} & \dots & n_{1m} \\
n_2 & 0 & n_2 & \dots & 0 & n_{21} & n_{22} & \dots & n_{2m} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
n_m & 0 & 0 & \dots & n_m & n_{m,1} & n_{m,2} & \dots & n_{m,m} \\
\sum_i n_{i1} & n_{11} & n_{12} & \dots & n_{1,m} & n_{11}^{(2)} & n_{12}^{(2)} & \dots & n_{1m}^{(2)} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\sum_i n_{im} & n_{m1} & n_{m2} & \dots & n_{m,m} & n_{m1}^{(2)} & n_{m2}^{(2)} & \dots & n_{mm}^{(2)}
\end{pmatrix}. \tag{7}
$$

We find the optimal design by maximising some optimality function of this matrix (see section below) subject to some constraints (such as $\sum_{i=1}^m n_i = n$, and other constraints particular to the network). Algorithms for finding optimal designs are considered in section 6.

3.1 Optimality Criteria

We seek an optimal design, which we define as minimising the variance of some function of the parameter estimates in our model described in equation (1). The optimal design is therefore equivalent to a particular function of the covariance matrix of the parameter estimates. For any possible design $X \in \mathcal{X}$ we could reasonably seek to minimise the average variance

1. of all pairwise differences of treatment effects,

$$
\frac{2}{m(m-1)} \sum_{j=1}^{m-1} \sum_{l=j+1}^{m} \mathrm{Var}(\widehat{\tau_j - \tau_l}).
$$

   This is defined to be A_s-optimality for estimating the differences in the treatment effects. We call this criterion function ϕ_1.

2. of all pairwise differences of network effects, γ_j.

   This is defined to be A_s-optimality for estimating the differences in the network effects. We call this criterion function ϕ_2.

Let $\boldsymbol{c}(a, b)$ be a length $2m+1$ vector which is a vector of zeroes, except the a-th element is 1 and b-th element is -1. Recall that I is a $(2m+1) \times (2m+1)$ matrix, the first row/column corresponding to the mean μ, the next m elements to the subject effects $\boldsymbol{\tau}$, and the final m elements to the network effects $\boldsymbol{\gamma}$. Then the optimality criteria above are equivalent to minimizing over all possible designs $X \in \mathcal{X}$

1. $\phi_1 = \sum_{j=2}^{m} \sum_{l=j+1}^{m+1} \boldsymbol{c}^T(j, l) I^{-1} \boldsymbol{c}(j, l)$.

2. $\phi_2 = \sum_{j=m+2}^{2m} \sum_{l=j+1}^{2m+1} \boldsymbol{c}^T(j, l) I^{-1} \boldsymbol{c}(j, l)$.

Other criteria may be appropriate depending on the aims of the experiment (e.g. a weighted sum of criteria 1 and 2). We call the values of these optimised functions at the optimal design the optimal function values, and represent them as $\phi_1^* = \min_{X \in \mathcal{X}} \phi_1$, etc.

We consider primarily the first criterion in the examples presented in this paper. Note that if we have n_j subjects given treatment j, we could (ignoring the network effect) estimate τ_j with variance σ^2/n_j. Thus we can estimate each difference $\tau_j - \tau_l$ with variance $\sigma^2/n_j + \sigma^2/n_l$. If we have an equi-replicate design such that $n_j = n/m$ for all j, then the minimum possible variance for the subject effects is $2\sigma^2/n_j = 2\sigma^2 m/n$.

3.1.1 Heuristic interpretation of information matrix

We can interpret our A_s-optimality criteria above in order to gain some heuristic insight about which designs (allocations of treatments) are likely to be optimal. As our information matrix I defined in (7) represents the information we can gain about the experiment performed with design matrix F, we wish this matrix I in some senses to be as "large as possible". We can also see this by noting that in calculating our optimality criteria ϕ_1 and ϕ_2 we take a linear function of the inverse of I.

In general, $\det I$ is large and designs seem to have better optimality criterion function values when they have

- an approximately balanced number of subjects given each treatment;

- a large number of occasions where the same treatment is applied to two neighbours of a subject;

- and a small number of occasions where the same treatment is applied to neighbours.

Of course, these features can conflict, but this gives an idea as to which properties are important in a good design.

3.2 Estimability

In practical situations, we are interested in the (variance of) the difference between treatments: whether advert 1 is more effective than advert 2, for our first motivating example. As is common in this situation, we will always have to impose some additional constraint on our parameters $\beta = (\mu, \tau, \gamma)^T$ to ensure they can be estimated uniquely. Without loss of generality, for the rest of this paper we have assumed that $\tau_m = 0$.

Even with this assumption, not all networks allow all effects to be estimated; consider the network where each node is connected to every other node. Then if subject i is given treatment j the response is

$$\mathbb{E}\left(Y_i | t(i) = 1\right) = \mu + \tau_1 + (n_1 - 1)\gamma_1 + n_2\gamma_2,$$
$$\mathbb{E}\left(Y_i | t(i) = 2\right) = \mu + \tau_2 + n_1\gamma_1 + (n_2 - 1)\gamma_2.$$

As $\tau_2 = 0$, we have only two different equations with which to estimate four unknowns (the mean μ, subject effect τ_1 and network effects γ_1 and γ_2,) and thus these are not estimable. Less extreme examples exist, with a number of edges less than the maximum possible.

In general for a linear model, $L\beta$ is estimable if and only if there is a K such that $L = KF$. A special case for all elements of β to be estimable, where L is the identity matrix, implies that K must be a generalised inverse F^- of F. A sufficient condition for this is that $(F^T F)^{-1} = I^{-1}$ exists such that $F^- = (F^T F)^{-1} F^T$, i.e. that I can be inverted, hence $\det(I) \neq 0$.

From the definition of I in equation (7), it is not clear what properties of a given network structure combined with a design lead to estimability; we can show trivially that n_i must be greater than zero (that all treatments must be given to at least one subject) otherwise a row and column of I will be zero, but further general statements applicable to all networks may not be possible. All the networks described in this paper are estimable with the sole assumption that $\tau_m = 0$, but further assumptions may have to be made for particular networks.
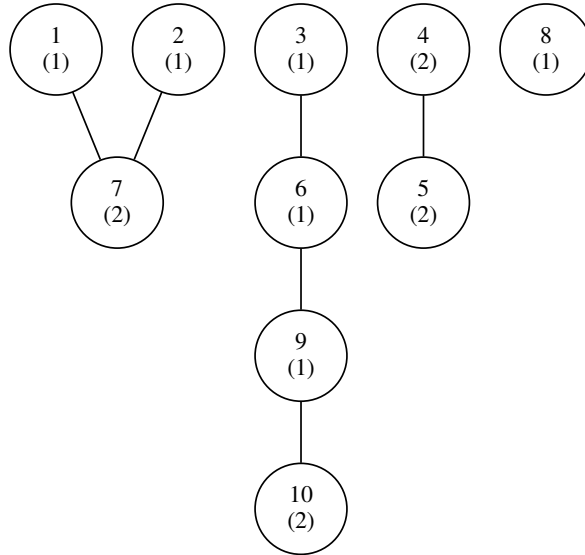
3.3 Example 1

Figure 1: An example network: optimal design for estimating the subject effect τ_1 for two treatments. Node numbers are shown above with treatment allocation for optimal design shown below in brackets

Let us suppose we have a network of 10 subjects, where some subjects are related to other subjects as shown in Figure 1. For example, subject 7 is connected to subjects 1 and 2 only, but subject 8 is not connected to any other subject.

By exhaustive search over all possible designs, we find that the optimal design for the $m = 2$ treatment case is $\{1, 1, 1, 2, 2, 1, 2, 1, 1, 1\}$ for optimality in estimating τ_1 via criterion ϕ_1 (i.e. we give treatment 1 to subjects 1,2,6,8,9, and 10 and treatment 2 to the other subjects). This design is that shown on Figure 1, and has optimal criterion value $\phi_1^* = 0.4186$. The design $\{1, 1, 1, 2, 2, 2, 1, 1, 1, 2\}$ is A_s-optimal for estimating the difference in the network effects $\gamma_1 - \gamma_2$ via criterion ϕ_2, with optimal criterion value $\phi_2^* = 0.2369$.

The optimal designs are different depending on what we wish to measure, with the result that the experimenter must think carefully about their experimental aims and the shape of the network before starting their experiment. This is a standard feature of optimal design.

The optimal designs are unbalanced; i.e treatments 1 and 2 are not applied to an equal number of subjects. This is an unusual property in optimal design.

The design which is optimal for ϕ_2 is very close to optimal for ϕ_1, with criterion value 0.4187. The design which is optimal for ϕ_1 does not do so well for the ϕ_2 criterion, with criterion value 0.4000.
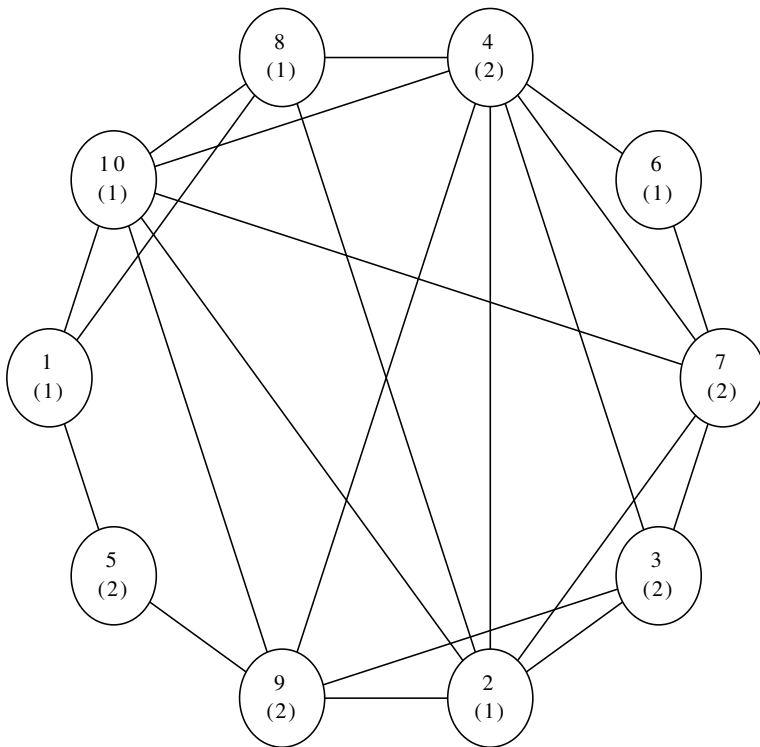
3.4 Example 2

Figure 2: A second example network: optimal design for estimating the subject effect τ_1 for two treatments. Node numbers are shown above with treatment allocation for optimal design shown below in brackets

We demonstrate another connectivity graph in Figure 2. This network again has 10 subjects connected by a friendship relationship, but the subjects are more likely to be connected to other subjects than in the previous network. Design $\{1, 1, 2, 2, 2, 1, 2, 1, 2, 1\}$ is A_s-optimal for estimating τ_1. This design is shown on Figure 2, with $\phi_1^* = 0.406$. Design $\{1, 2, 1, 2, 1, 2, 2, 1, 2, 1\}$ is optimal for estimating the difference in network effects $\gamma_2 - \gamma_1$, with $\phi_2^* = 0.1257$.

Again, the designs which are optimal for ϕ_1 and ϕ_2 are different. However for this network the optimal design for both criteria are balanced, such that 5 subjects are given each treatment,

11

but that for estimating the network effects (ϕ_2) is not.

The design which is optimal for ϕ_2 has criterion value for ϕ_1 of 0.4643. The design which is optimal for ϕ_1 has criterion value for ϕ_2 of 0.2073.

4 Misspecification of model: Efficiencies and biases

We now consider whether we need to take network effects into account at the design stage for an experiment. It may be that designs which do not consider a network effect are as good or only a little worse than designs that do; in this section we show that ignoring a network effect can lead to very poor designs in terms of low efficiency (high variance) and/or bias in our estimators.

4.1 Simulation to find efficiencies of designs

Let us now suppose we design an experiment in which we wrongly assume that there is no network effect, but the network effect acts according to the linear network effects model described above in equation (1).
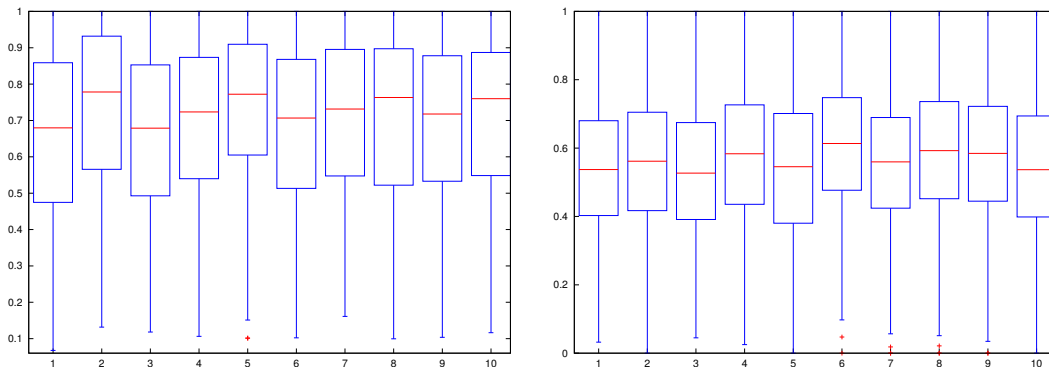

Figure 3: Efficiencies of balanced designs ignoring network effects. Efficiency (vertical axis) for 10 random networks of 12 nodes with probability of connection 0.6 and (left) 2 and (right) 3 treatments. Each boxplot represents all possible balanced designs on a given social network.

We consider ϕ_1 here. For $n = 12$ subjects and both $m = 2$ and $m = 3$ treatments, we generate 10 Erdős-Renyi random networks where two nodes are connected fixed probability $p = 0.6$ and calculate the efficiencies of all possible balanced designs (there are $\frac{n!}{((n/m)!)^m}$ of these as m divides n, so 924 for $m = 2$ and 6930 for $m = 3$), compared to the optimal design for that random network if there is a network effect. The results are shown as Figure 3.

We see that, for instance, for the right plot in Figure 3 ($n = 12, m = 3, p = 0.6$) that the median efficiency of the set of balanced designs is typically around 0.6. For all networks we can find a balanced design with very bad efficiency (for $m = 3$ network 5, for example, has a balanced design with efficiency of 0, where all subject effects are not estimable). The range and interquartile range of efficiencies are always large; we conclude that if there is a network effect, it is vital to include it for an efficient design.
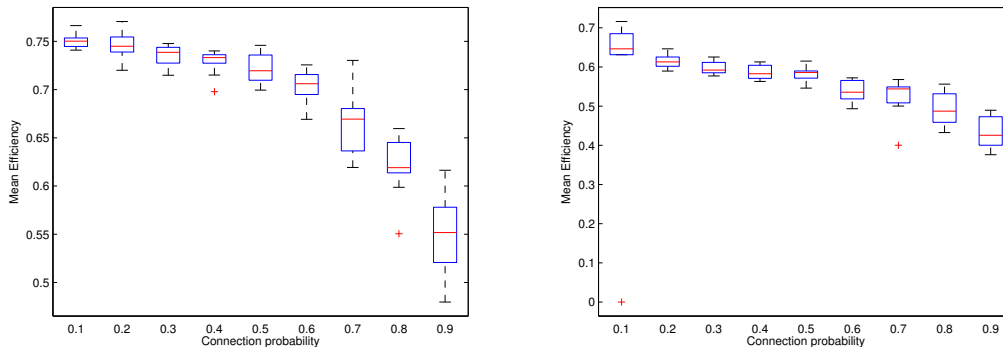
Figure 4: Efficiencies of balanced designs ignoring network effects. Efficiency for random networks of 12 nodes with varying probability of connection for (left) 2 and (right) 3 treatments. Each boxplot represents 10 randomly chosen social networks.

We plot in Figure 4 how the mean efficiency varies with connection probability. Each boxplot now represents mean efficiencies on each of 10 randomly generated Erdős-Renyi networks with fixed connection probability. We see that as the connection probability increases, the efficiency decreases; also the efficiencies of the balanced design for $m = 3$ treatments are less than those for $m = 2$, confirming the pattern seen for $p = 0.6$. We see that one of the networks with $p = 0.1$ has mean efficiency of 0, which means that there are no balanced designs which allow all parameters to be estimated.

4.1.1 Example 1 revisited

Consider example 1 (see Figure 1), where $m = 2$. If we believed there is no network effect in the model, we would conclude that balanced designs with 5 subjects chosen at random given each of treatments 1 and 2 were optimal. We might chose an arbitrary balanced design $\{1, 1, 1, 1, 1, 2, 2, 2, 2, 2\}$. The design for estimating the subject effects using model (1) with this design has $\phi_1 = 0.4502$. We saw that the optimal design had value $\phi_1^* = 0.4186$. The efficiency of this arbitrary balanced design is thus $0.4180/0.4502 = 92.99\%$. The average variance of estimates for τ_1 is slightly bigger than with the optimal design when the network effect is accounted for.

4.2 Bias of design

Now let us consider the bias of the design. If we know the adjacency matrix, and assume that there is a network effect $\boldsymbol{\gamma} \neq 0$, such that the true model is as specified in equation (1), we can find the optimal design matrix X_C by exhaustive search as above and hence form the extended design matrix $F_C = \begin{pmatrix} \mathbf{1} & X^* & AX \end{pmatrix}$, where X^* is the design matrix X without its last column to impose the constraint $\tau_m = 0$. From the Gauss-Markov Theorem, we find a best linear unbiased estimator for $\boldsymbol{\beta}$ under these assumptions as $\hat{\boldsymbol{\beta}}_C = (F_C^T F_C)^{-1} F_C^T Y$.

13

If we perform an experiment where we wrongly assume there are no network effects, then our model is $\boldsymbol{Y} = F_R(\mu \quad \boldsymbol{\tau}^*)^T + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 J_n)$, with J_n being the $n \times n$ identity matrix; then the least squares estimator of the mean, subject, and network effects $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}_R = [(F_R^T F_R)^{-1} F_R^T \boldsymbol{Y} \quad \boldsymbol{0}_m]^T$, where the extended design matrix for any design which we would believe to be optimal for estimating the subject effects with no network effect can be found as $F_R = \begin{pmatrix} \boldsymbol{1} & X_R^* \end{pmatrix}$. Here X_R^* is defined as X_R with the last column deleted to impose the constraint $\tau_m = 0$. The biases in our estimators, under the assumption there is a network effect and we do not account for it, is $\mathbb{E}(\widehat{\boldsymbol{\beta}_R}) - \boldsymbol{\beta} = \mathbb{E}(\widehat{\boldsymbol{\beta}_R}) - \mathbb{E}(\widehat{\boldsymbol{\beta}_C}) = \mathbb{E}(\widehat{\boldsymbol{\beta}_R} - \widehat{\boldsymbol{\beta}_C})$ as $\widehat{\boldsymbol{\beta}_C}$ is an unbiased estimator. We can thus find the bias as

$$
\begin{aligned}
\mathbb{E}(\widehat{\boldsymbol{\beta}_R} - \widehat{\boldsymbol{\beta}_C}) &= \mathbb{E}\left[\begin{pmatrix} (F_R^T F_R)^{-1} F_R^T \\ \boldsymbol{0}_{m \times n} \end{pmatrix} \boldsymbol{Y} - (F_C^T F_C)^{-1} F_C^T \boldsymbol{Y}\right] \\
&= \left[\begin{pmatrix} (F_R^T F_R)^{-1} F_R^T \\ \boldsymbol{0}_{m \times n} \end{pmatrix} - (F_C^T F_C)^{-1} F_C^T\right] \mathbb{E}[\boldsymbol{Y}] \\
&= \left[\begin{pmatrix} (F_R^T F_R)^{-1} F_R^T \\ \boldsymbol{0}_{m \times n} \end{pmatrix} - (F_C^T F_C)^{-1} F_C^T\right] F_C \beta,
\end{aligned}
$$

which is not in general zero, and neither are the elements corresponding to just the subject effects (elements 2 to m of this expectation vector). In general, the bias of any contrast of an estimate of a subject effect $\hat{\tau}_j - \hat{\tau}_l$ will also not be zero, and thus ignoring the network effects in calculating a design will produce a biased estimator of the subject effects.

In examples 1 and 2, if we wrongly assumed there was no network effect when in fact it existed, then for the balanced design $\{1,1,1,1,1,2,2,2,2,2\}$ which is optimal in the case there were no network effects, we can calculate the bias in terms of unknown parameters $\boldsymbol{\beta}$ for examples 1 and 2 as

$$
\text{Bias(Ex 1)} = \begin{pmatrix} 0 & 0 & 0.6 & 0.8 \\ 0 & 0 & -0.2 & -0.2 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \gamma_1 \\ \gamma_2 \end{pmatrix}, \qquad \text{Bias(Ex 2)} = \begin{pmatrix} 0 & 0 & 2.8 & 1.6 \\ 0 & 0 & -1.2 & 1.2 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \gamma_1 \\ \gamma_2 \end{pmatrix}.
$$

We of course see that the biases in the network effects (γ_1 and γ_2) are equal to the negative of the effects themselves, as we have ignored them; there is no bias in the estimate of τ_2 as we have assumed this to be zero and it is not included in $\boldsymbol{\beta}$. The biases for μ and τ_1 are (different) linear combinations of the unknown values of the network effects.

Our main interest will often be in the (difference in) treatment effect. In Example 1 we find $\text{Bias}(\tau_1) = -0.2\gamma_1 - 0.2\gamma_2$. If the true value of the network effects are zero (such that $\gamma_1 = \gamma_2 = 0$) then our design will produce unbiased estimators. If our network effects have similar magnitudes, but of opposite directions, the bias will be close to zero.

Except in particular cases, which vary across different networks and are likely to be outside the control of the experimentor, the network effects are substantial compared to the subject effects, ignoring network effects might generally lead to the subject effects being estimated with very large bias. Thus by not taking into account a network effect in our design, we produce an experiment which has higher variance than necessary, and/or biased estimators.

4.3 Misspecification of network

We are also interested in the case where the network is misspecified- that is we do not know with certainty whether there is a link between two subjects. We simulate this situation by again considering Example 2 for a social network. For any two nodes in the network for example 2 given by adjacency matrix A, we flip $A_{ij} = A_{ji}$ from $1 \rightarrow 0$ with probability p, and from $0 \rightarrow 1$ with the same probability. We calculate the efficiency of using the optimal design for the mispecified network when the network is really as described in example 2, and do this for 1000 simulated networks.
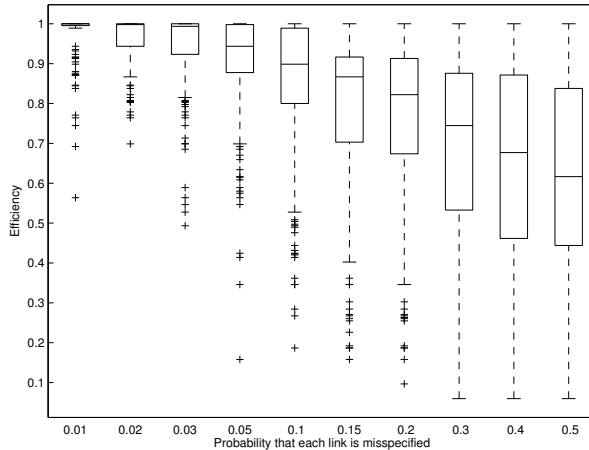

Figure 5: Efficiencies of misspecified networks. Efficiency (y-axis) for 1000 random networks with probability that each link is misspecified given on x-axis.

We see in Figure 5 that efficiency tends to decrease with increasing probability of misspecification. There is a large variation in efficiencies even for small probabilities of misspecification, but the typical (median) design is generally reasonably efficient. In practice, simulations such as this could help experimenters to find designs robust against misspecification.

5 Effect of network structure on optimal design

We consider how the macroscopic properties of the network affect the design. One way of assessing the macroscopic properties of the graph is through the graph's spectrum, which is the set of eigenvalues of the adjacency matrix A. The Laplacian spectrum is also often studied, i.e the eigenvalues of the Laplacian L where $L = D - A$. The degree matrix D specifies the number of neighbours each node has: $D_{ii} = \sum_j A_{ij}$, $D_{ij} = 0$ for $i \neq j$.

These eigenvalues have various interpretations and summarise important macroscopic properties of the graph; for example, the second smallest eigenvalue of the Laplacian matrix is the algebraic connectivity of our network. This gives a measure of how densely connected the graph is.

We wish to test the belief that the macroscopic properties of the network to the optimal design for that network. We do this through two examples.

Algebraic	Design									
Connectivity	A	B	C	D	E	F	G	H	I	J
3.6977	7	2	9	5	**1**	3	8	4	10	6
3.7046	5	9	7	6	10	2	3	8	**1**	4
4.3028	8	7	2	5	6	9	1	10	3	4
4.3174	8	3	7	4	5	6	9	2	**1**	10
4.5392	3	3	6	7	**1**	1	8	3	9	9
4.5494	9	5	2	7	3	10	8	4	6	1
4.8631	10	4	3	6	**1**	8	2	5	9	6
5.3028	6	4	**1**	3	2	9	7	10	5	8

Table 1: For 8 different networks, with different algebraic connectivities, 10 different designs denoted A to J are ranked from best (1) to worst (10). A better design has lower ϕ_1. Note that two designs are equally good, they both take the higher rank.

5.1 Example: Graph Spectra

We test the hypothesis that the algebraic connectivity, calculated through the spectra, affects the optimal design for that network. To test this hypothesis we formed seven Erdős-Renyi random graphs with $n = 6$ nodes, where each pair of nodes is connected with equal probability 0.5. We calculated the spectrum and the Laplacian spectrum.

We tabulate the results for this as Table 1. Each row of the table corresponds to one of eight different networks. For each of ten (randomly chosen) designs labelled A to J, we rank the designs in terms of the best (lowest) value of ϕ_1, such that 1 indicates the best design for that network, and 10 the worst design. Consider, for example Design F. It is the best of the ten designs for the network with algebraic connectivity 4.5392, but worse for that network with algebraic connectivity only slightly higher at 4.5494.

We see that each different design may be good, bad, or indifferent for each particular network. There seems to be no fixed ordering, such that designs tend to be better for all networks. There seems to be no correlation between the quality of a design and the algebraic connectivity.

We see similar patterns for other eigenvalues and Laplacian eigenvalues, but omit them for conciseness and as their physical interpretation is less clear.

5.2 Co-spectral Graphs

As a further example to show that macroscopic properties of a network do not determine the design, consider the two graphs pictured as Figure 6. These graphs are clearly very different, although they have the same spectrum (-2,0,0,0,2). However, the first graph has an optimal design with A_s-optimality function value 2; in fact, of the 16 distinct designs for $n = 5$ subjects, 8 of them are optimal, and 8 are sub-optimal, and in this case lead to inestimability of the subject effects. For the second graph all designs lead to an inestimable difference in subject effects; i.e. there is no optimal design, or rather all designs are equally ineffective.
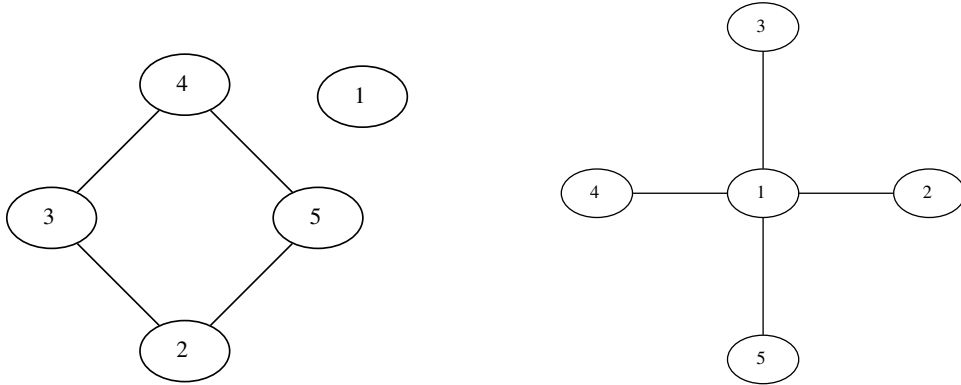
Figure 6: Two co-spectral graphs

Wilson and Zhu (2008) consider graphs which are co-spectral in their adjacency matrix, Laplacian, as well as two other criteria (sign-less Laplacian, and normalised Laplacian). We compare the pair with upper-right triangle matrices of A

$$01111101010101011010000100010111001101011011$$
$$\text{and } 11010010111110000001101101101010100111001111010} \qquad .$$

For example, with $m = 3$, the ϕ_1^* are 0.6263 and 0.6274 respectively. The ϕ_2^* are also different (1.158 and 1.350 respectively), with different optimal designs for the two co-spectral networks. We again see that knowing the spectrum or Laplacian spectrum does not allow us to determine which design is optimal without also knowing the detailed network structure.

5.3 Summary of relationship between network properties and optimal design

We wanted to test the hypothesis that there is a relationship between the large-scale properties of the network and the optimal design for that network; as many of the large-scale properties of the networks are encoded in the spectrum of the graph (and e.g. Laplacian spectrum), it is surprising that two graphs which are co-spectral produce very different designs. Thus we can be sure that the large-scale properties alone of the graph do not determine the optimal design.

It may be that the large-scale properties of a graph do not determine the design absolutely, but there is an association between these properties and the design; Table 1 shows clearly that even simpler properties, such as algebraic connectivity, do not have a strong effect on determining the design. Thus for any network, it is important to know the connectivity structure of the graph exactly in order to be confident of a good design, and we cannot hope to find a good algorithm for finding design simply by knowing large scale properties of the network.

For some online social networks, such as Twitter, the structure of the network is publicly available, whereas for others, such as Facebook, user privacy prevents anyone but an administrator knowing the global connectivity structure of the network. For less well defined social networks, such as friendship networks, there may be uncertainty in a connection. Finally,

the strength of connections on a network (corresponding to a non-binary matrix A) may also strongly affect the design. Future work will include generalising this work to the case where we do not know the complete network structure, but have some information on it.

6 Algorithms for finding optimal design

Recall from section 3.1 that we wish to find an optimal design by finding the best design (an allocation of treatments to experimental units) to minimise some (linear) function of our information matrix, I, defined in equation (7). The criteria ϕ_1 and ϕ_2 contained in that section define reasonable optimality criterion.

As the information matrix I depends on some summary statistics of the design, we might consider a constrained optimisation to find the optimal design. However, in all except trivial networks, the function to be optimised is complicated (even for $m = 2$ it is a quartic) and, computationally, we consider other methods for finding the optimal design.

For small networks, we exhaustively consider each design within the design space $\mathcal{X}$, calculate its Fisher information matrix, and hence the value of the design criterion. The design which yields the best value of the design criterion function is our optimal design for that criterion. The exhaustive search is computationally expensive; $|\mathcal{X}| = m^n$, which becomes unmanageable for moderate network sizes n or treatments m. We can reduce the size of the search region by two methods: considering symmetry of labels in the design, and symmetries in the graph.

6.1 Symmetry of labels

For our criteria we are only interested in differences between treatments, the treatment effects themselves are irrelevant, and treatments are equivalent up to relabelling. For example, 1,2,2,3,1 and 1,3,3,2,1 are equivalent. We can thus reorder any design so that we only evaluate designs where the first occurrence of label j must come before the first occurrence of label $j + 1$. Without loss of generality, we assign treatment 1 to experimental unit 1.

6.2 Symmetry of graphs

In Figure 7, subgraphs 1 and 2 are exchangeable; i.e if we consider 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 thus reduce our design space greatly if we can identify subgraphs where the designs are exchangeable. 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 known as the Graph Automorphism Problem. Unfortunately, it is not known whether this problem is NP-complete or not (Lubiw 1981), and no general algorithm for reducing the design space is available.

In practice, we suggest that, except in the $m = 2$ case, it is generally quicker computationally to find graph automorphisms to reduce the design space before performing an exhaustive search.
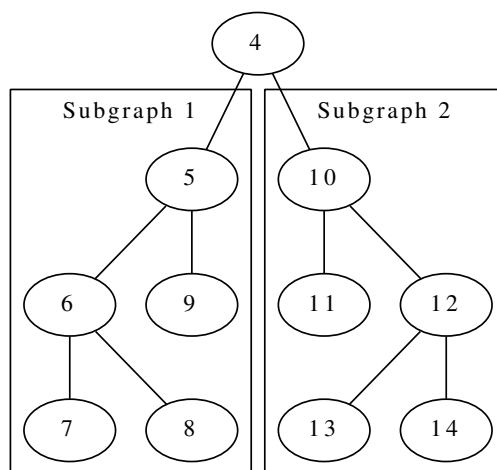
Figure 7: Demonstration of automorphism.

6.3 Non-exhaustive search

For large networks, exhaustive search may be difficult due to the number of possible designs ($|\mathcal{X}| = m^n$). We have constructed an algorithm (a type of Fedorov exchange algorithm inspired by Fedorov (1972)) to find an efficient design which may be near optimal.

1. Pick a random balanced design and evaluate the information function for that design.

2. Either

    - with probability $p = 0.9$: swap two randomly chosen labels
    - otherwise pick a label randomly and relabel it with a randomly chosen label.

3. Evaluate the information function for the new design;

    - if it is smaller (better), accept the new design.
    - if it is larger (worse), accept the new design with small probability $\delta = 0.05$.

4. Repeat from step 2 until a fixed number of iterations are completed.

We find 10,000 iterations to be effective for the examples presented here. Step 3 ensures that we can sometimes pick a locally worse design in the hope of eventually finding a globally optimal design.

Extending this simple algorithm is an area for future work; for example simulated annealing algorithms for finding optimal designs such as described by Wit et al. (2005) would also an effective solution to avoiding finding locally optimal designs which are not globally optimal.

7 Further examples

We present further examples to show how this method can find optimal designs for a wide class of experiments where the treatment applied to a subject affects other subjects.

7.1 Example 3: A Social Network

Suppose we design an experiment to determine the effectiveness of a new advertisement. Ten people are sent a video advertisement for a type of soft drink (treatment 1), and ten are shown no advertising (treatment 2). The amount of soft drink each person buys in the following week is then monitored. The marketing executives are interested not only in how effective the advertisement is, but also how effective it is at being conveyed to friends of those in the group which originally saw the advertisement (this is known as viral marketing), so 20 people are chosen such that the connectivity structure of an online social network of the subjects is known.
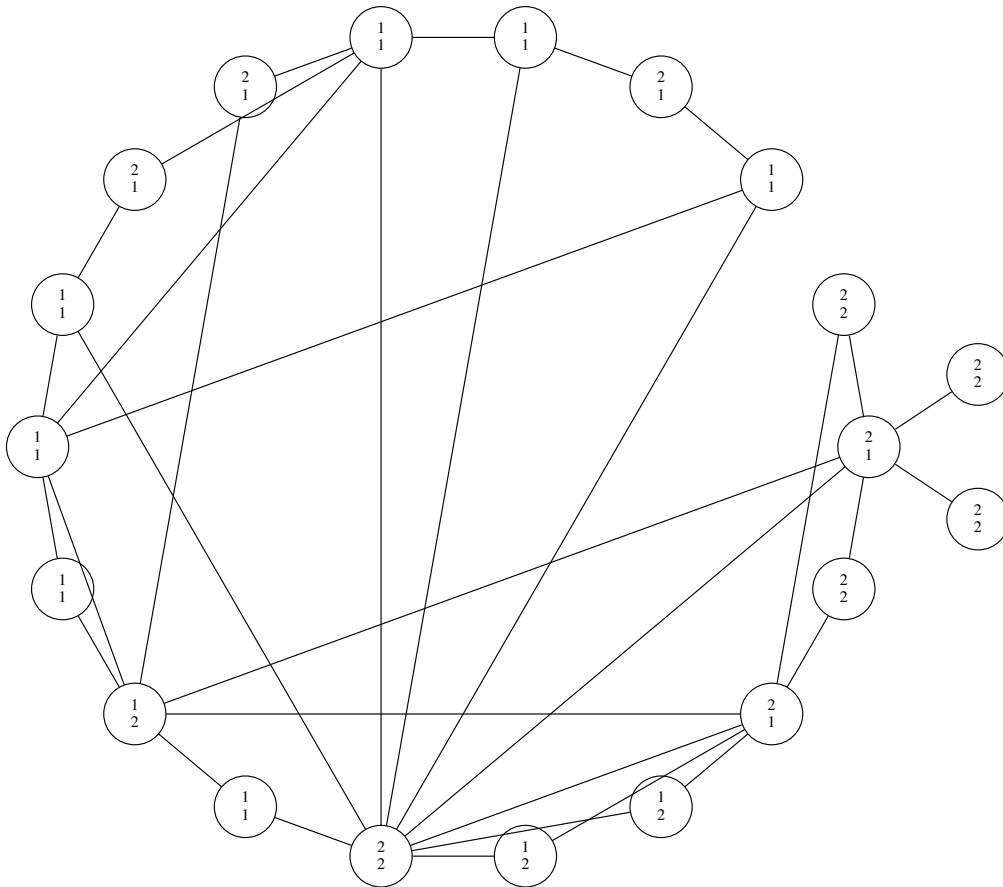

Figure 8: Optimal design for estimating subject(top) and network(bottom) effects of an advert

We see in Figure 8 the shape of the social network for these people. The top figure for each

node presents the optimal design (for ϕ_1) for estimating (with minimum variance) how much extra soft drink people buy as a direct result of having been sent the advert. If our linear network effects model is true, we find $\phi_1^* = 0.2002$. The bottom number in each node shows the optimal design for ϕ_2 for estimating the network effect (the viral effect of the advert), with $\phi_2^* = 0.0284$. This design is also not bad for estimating the subject effects (variance $0.2007\sigma^2$).

7.2 Example 4: Block Designs for left and right treatment effects

Druilhet (1999) shows that this example design is optimal for estimating $m = 4$ treatment effects:

$$\begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix} \begin{pmatrix} 2 & 3 & 4 \\ 1 & 4 & 3 \\ 4 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 4 \\ 3 \end{pmatrix}$$

Each number represents a treatment applied to a plot, and the treatment will affect its left and right neighbours within the block, each block being a horizontal line of plots. There is a block effect for all subjects in a block. The plots in the left and right brackets are "guard" plots, and no response is gathered from these plots.
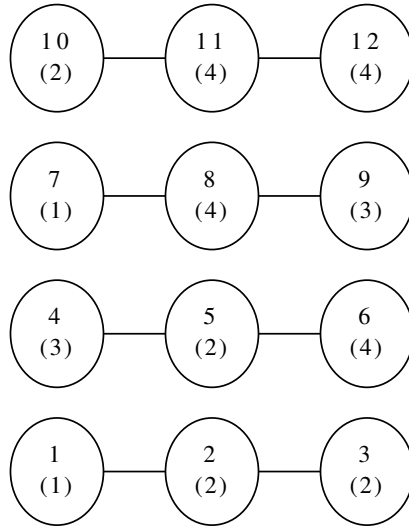

Figure 9: A blocked experiment with left and right treatment effects. Top numbers are plot labels (nodes in the network) and bottom numbers the optimal allocation for $m = 4$ treatments.

We represent the plots as a network with a connectivity structure as shown in Figure 9. We also plot the $m = 4$ optimal design for criterion 1 on that figure. Without guard plots, it is interesting to note that the optimal design is not balanced, let alone neighbour balanced; for instance, treatment 2 appears 4 times, but treatment 1 only twice. However, Druilhet's model assumes a block effect which we do not have in our present model, so a direct comparison may not be fair; with our method we can design an experiment without guard plots.

7.3 Example 5: Non-rectangular field trials

Most field trial experiments focus on rectangular fields. Let us assume we have an irregularly shaped field divided into plots as shown in Table 2. An integer represents the number of a

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

Table 2: An irregularly shaped collection of plots in a field.

plot. We wish to perform a trial on different fertilisers, to estimate the difference between fertilizer (treatment) effects with minimum average variance. We assume that any fertiliser applied might leach to any of the 8 plots that touch the treated plot, including those only touching at corners.

Here, we can represent the connectivity structure by a network as shown in Figure 10, together with the optimal design for ϕ_1. This design is balanced, with all treatments equally replicated. Here, $\phi_1^* = 0.4052$. The minimum average variance possible with no restriction on comparisons $2 \times \frac{\sigma^2 m}{n} = 0.4\sigma^2$. It is interesting to note how close we can get to the variance for the unrestricted case with this unusual shape of field.

7.4 Example 6: Extension to a crossover trial with planned dropouts

Usually in a crossover trial, including a dropout in the design could be tricky. Let us assume we are performing a crossover trial on four subjects (a, b, c, and d) over four periods with three treatments (labelled 1 to 3), each of which is believed to have a wash-out time of one period. Assume further that it becomes clear that participant b will not be able to take the treatment in the third period of the trial. This is represented by Table 3.

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

Table 3: A crossover trial. Left: Experimental units numbered as subject/period combinations. Right: Optimal design.

We use each subject/period combination as a node in our network, and we modify our methodology such that links are uni-directional by relaxing the restriction that A is symmetric; our Linear Network Effects Model in Equation (1) is still valid, and we perform an exhaustive search as before to minimise ϕ_1.
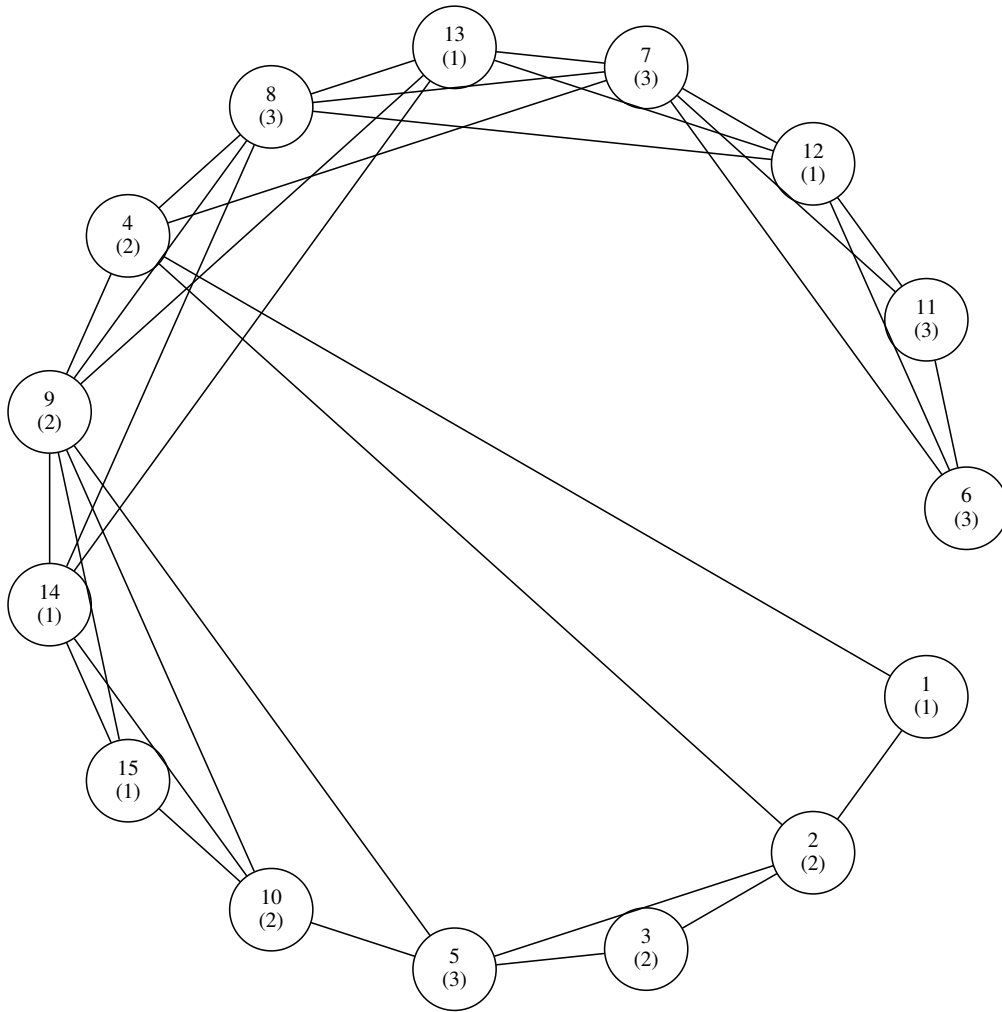
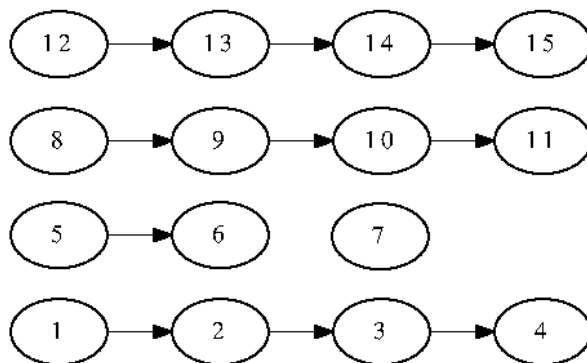Figure 10: An experiment on an irregularly shaped field.

Figure 11: Cross-over trial with planned dropout. Nodes in network as subject/period combinations.

The network for this experiment is shown as Figure 11. The optimal design found by considering this experiment as a network is presented as Table 3. The optimal design has $\phi_1^* = 0.4128$. Note that it is quite different in form from usual cross-over designs.

7.5 Example 7: Motivating example revisited

Recall that we assume we have a network of 10000 customers, and want to assess the effectiveness of sending 4 different adverts each to a subset of 25 customers, so that 100 customers in total were selected. After the experiment, the best advertisement would be sent to the rest of the customers. An advertisement is deemed effective if more customers click on a hyperlink in a message. It can be determined which customers click on a hyperlink. Although in this example 9900 customers are not directly sent an advert, they may see a test advert sent to one of the subsets (perhaps one of the customers sends to them or shares it publicly) and respond, so all customers are measured.

Marketers on many online social networks (e.g. Facebook and Twitter) pay per "impression", for each time that a message is displayed to a user. It is common in marketing to assess the effectiveness of an advert on a small group of people, before paying for it to be promoted to a much larger network. Each user may be exposed to their assigned advert many times.

To generate a typical network of 10,000 customers; we start with two linked nodes and add 9998 further nodes individually according to a preferential attachment (Barabási-Albert) model, such that each node adds two links with probability of attachment to an existing node proportional to the degree of the existing node.

We then wish to assess designs with four adverts. We use five treatments: treatments 1 to 4 are coded to mean that adverts 1 to 4 are sent to a customer, and we invent a special treatment 0 which means that the customer is not selected. It is important that we include customers that are not given an advertisement in our experiment, as even though we do not show them an advertisement directly, the advertisement may be passed to them virally and they may click on the hyperlink.

As our network of $n = 10000$ does not allow us to perform an exhaustive search, we compare two methods: 1) the lowest information function found from a number of iterations of the exchange algorithm presented in section 6.3; 2) the lowest information function found after evaluating a number of randomly chosen designs.

The results are presented as Figure 12 by a plot of the lowest (best) information function value for both ϕ_1 and ϕ_2 against the number of designs evaluated for each algorithm, on a log-log scale.

For both algorithms, the algorithm converges to find designs with function values close to optimal. In this example, choosing random designs seems to work better than an exchange algorithm, although the difference in efficiencies between the two designs finally found (noting the log scale) is minimal. The results are similar for both ϕ_1 and ϕ_2.

We investigated for this example whether there was any correlation between the degree of a node and the probability of receiving any treatment. We found the linear correlation to be 0.03 for ϕ_1 and 0.0381 for ϕ_2. Performing a t-test showed that this correlation was significant but tiny at the 1% level of significance. It is clear that for our network effect, by giving a treatment to a highly connected node, we will be able to measure the effect more
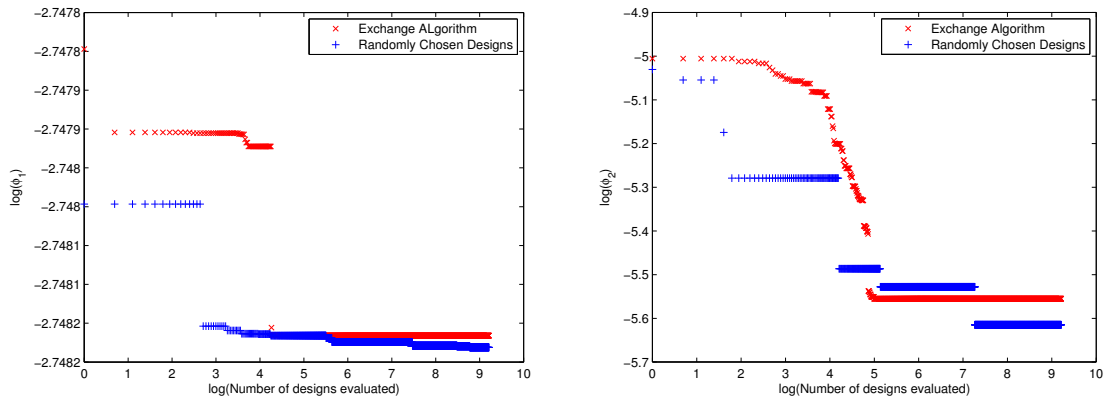
Figure 12: Comparison of algorithms for optimal design of large networks. Information function value against number of designs evaluated on log-log scale. Exchange algorithm in red and random selection of designs in blue. Left: ϕ_1. Right: ϕ_2.

effectively as the network effect will propagate to all neighbours of the highly connected node. For calculating the subject effect, it is also advantageous to select highly connected nodes as a highly connected node will experience many more network effects from its neighbours, which will allow the network effect to be estimated more precisely, which in turn will allow the subject effect of the treatment applied to the highly connected node to be estimated more precisely.

8   Conclusions

We have presented a new methodology for designing experiments on social networks, an area which has seen little research. We introduced a linear network effects model that encompasses a wide class of experiments where the treatment on one subject affects the other subjects. We have shown how one can readily find optimal designs for these experiments, and described some techniques to make this exhaustive search procedure faster.

We have shown the problems in inflated bias and variance that may be experienced when we neglect a network effect in an experiment where one is present. Finally, we have presented examples from a wide area of experiments to demonstrate the flexibility of our method, focusing on marketing experiments on social networks, but also showing examples on agricultural and crossover trials. The method seems to be widely applicable and simple to implement, and we hope this paper serves as an introduction to a very wide class of models which can benefit from experimental design techniques.

Acknowledgment

reading of the paper, which resulted in many improvements, and for checking the optimal designs.

References

Azaïs, J.-M. and Druilhet, P. (1997). Optimality of neighbour balanced designs when neighbour effects are neglected. *Journal of Statistical Planning and Inference*, 64(2):353–367.

Bailey, R. (2008). *Design of Comparative Experiments*. Cambridge University Press.

Besag, J. and Kempton, R. (1986). Statistical analysis of field experiments using neighbouring plots. *Biometrics*, 42(2):231–251.

Doreian, P., Freeman, L., White, D., and Romney, A. (1989). Models of network effects on social actors. *Research Methods in Social Network Analysis*, pages 295–317.

Druilhet, P. (1999). Optimality of neighbour balanced designs. *Journal of Statistical Planning and Inference*, 81(1):141–152.

Fedorov, V. V. (1972). *Theory of optimal experiments*. Elsevier.

Hedayat, A. and Zhao, W. (1990). Optimal two-period repeated measurements designs. *The Annals of Statistics*, pages 1805–1816.

Jones, B. and Kenward, M. G. (2003). *Design and Analysis of Cross-Over Trials*. Chapman & Hall.

Kunert, J. and Martin, R. J. (2000). On the determination of optimal designs for an interference model. *The Annals of Statistics*, 28(6):1728–1742.

Kunert, J. and Mersmann, S. (2011). Optimal designs for an interference model. *Journal of Statistical Planning and Inference*, 141(4):1623–1632.

Leenders, R. (2002). Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks*, 24(1):21–47.

Lubiw, A. (1981). Some NP-complete problems similar to graph isomorphism. *SIAM Journal on Computing*, 10:11.

McConnell, M., Sinclair, B., and Green, D. P. (2010). Detecting social networks: design and analysis of multilevel experiments. In *Third Annual Center for Experimental Social Science and New York University Experimental Political Science Conference*.

Pearce, S. C. (1957). Experimenting with organisms as blocks. *Biometrika*, 44(1/2):141–149.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 6(1):34–58.

Wilson, R. C. and Zhu, P. (2008). A study of graph spectra for comparing graphs and trees. *Pattern Recognition*, 41(9):2833–2841.

Wit, E., Nobile, A., and Khanin, R. (2005). Near-optimal designs for dual channel microarray studies. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(5):817–830.