Design of Experiments for Learning from Online Networks (4-Week Review)

Given the increasing reliance on online advertising by commercial firms and the global marketing successes seen by brands such as Netflix [USK16] and Microsoft [KT17], there is a growing contemporary focus on the statistical underpinning of online experiments [Lar+24]. A particular type of such experiment is A/B testing: a method of comparing the success of two versions of an advertisement, product, or webpage by analysing the purchasing intent of its end users [You14]. Moreover, graph theory has seen a rise in popularity due to its adaptability to different disciplines [GLH14]. Despite the increasing volume of literature on both A/B testing [Qui+24] and graph theory [GLH14], there is relatively scarce research on how to conduct online experiments on a large network [PGS17]. This PhD project aims to synthesise ideas from these two emerging fields; we wish to establish a novel approach to conducting experiments over large networks. This four-week review document provides a preliminary plan to be able to achieve this goal.

1 Background

In this section, we provide an overview of the problem that this PhD project aims to solve.

1.1 Graphs and Networks

Networks are a useful tool to illustrate connections between people, processes, and data [Gos+18; GLH14].

Definition 1.1. (Network [PGS17]) An undirected network is a set of nodes, $\mathcal{N} = \{\nu_1, \nu_2, \ldots, \nu_N\}$, and a set of edges, $\mathcal{E} = \{e_1, e_2, \ldots, e_K\}$, for $K, N \in \mathbb{N}$ such that each edge in $\mathcal{E}$ connects an element in $\mathcal{N}$ to another (not necessarily distinct) element of $\mathcal{N}$.

Nodes may represent people and the edges joining them may take the form of family or friend connections [PGS17]. In the context of online A/B testing, one may use nodes to represent end users of a website and the corresponding edges may be used to represent the user's Facebook friend networks or LinkedIn connections [BSS23].

Understanding the structure of social interactions may yield interesting results from A/B testing; it has been documented that social influence may affect the success of advertising campaigns and response behaviour [Gra+15]. This will be discussed further in Section 1.3. Figure 1 demonstrates an example of a simple network.

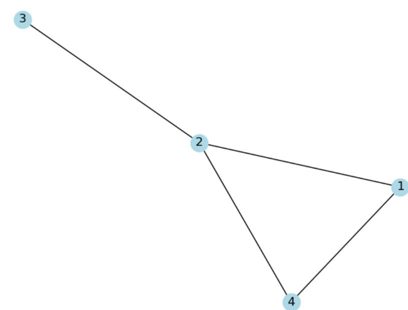

Figure 1: A network with node set $\mathcal{N} = \{1, 2, 3, 4\}$ and edge set $\mathcal{E} = \{(1, 2), (1, 4), (2, 4), (2, 3)\}$.

Moreover, one can describe the connectivity of the graph by writing its adjacency matrix.

Definition 1.2. (Adjacency Matrix [Bap10]) Given an undirected network (as in Definition 1.1), the adjacency matrix, A, is defined as a symmetric $N \times N$ matrix where:

$$A_{ij} = \begin{cases} 1 \text{ if } (i, j) \in \mathcal{E} \\ 0 \text{ if } (i, j) \notin \mathcal{E} \end{cases}$$

1.2 Experimental Design

In the context of online advertising, a marketer may wish to test the success of their advertisement campaign on a population by measuring click-through rates and engagement activity (i.e., comments, likes or shares), for example. In addition, the marketer may want to determine the success of their campaign between and across various demographic groups. This is, in fact, an online experiment and, with it, a hypothesis test to determine whether the new advert is more successful than a previous campaign. A solid understanding of experimental design enables one to test this hypothesis more definitively, by knowing which factors to isolate, which participants to group together, and how to ensure extraneous factors are not diminishing the statistical power of conclusions [Mon20].

Remark 1.3. Montgomery [Mon20] states that there are three fundamental elements to sound experimental design: blocking, randomisation, and replication:

1. **Blocking** refers to the practice of separating a sample into groups formed by units which are expected to behave similarly.
2. **Randomisation** is the process of allocating treatments and running order randomly. This is to ensure that each run is independent from one another. Such assumptions of independence are critical for ANOVA analyses.
3. **Replication** requires that we independently repeat runs of each treatment combination. A greater number of replicates enables the user to more accurately estimate experimental error. This is vitally important to determine to what extent error impacts the conclusion.

Mead et al. [MGM12] further state that these three fundamental principles of experimental design assist in maximising the statistical power of an experiment; they enable one to be more confident in rejecting null hypotheses based on experimental data. In order to demonstrate the foundations of experimental design, consider the following example.

Example 1.4. An online marketer wishes to amend an online advert sent to Britons of all ages, offering ten-percent off shoes, to state "10% OFF" rather than "10% DISCOUNT". The marketer believes the snappier wording will see a higher click-through rate. This could be seen as an example of an A/B test, as there is one factor (advert wording) with two levels ("OFF" and "DISCOUNT"). Mathematically, this looks like a hypothesis test with a null hypothesis, H_0, and an alternate hypothesis, H_1, comparing the average click through rates, μ_a for the wording "OFF" ($a = 1$) and the wording "DISCOUNT" ($a = 2$):

$$H_0 : \ \mu_1 = \mu_2$$
$$H_1 : \ \mu_1 \neq \mu_2$$

In order to implement the first principle outlined in Remark 1.3, the marketer ensures participants are **blocked** by age group. This is because the marketer has prior knowledge that each age group responds similarly within their group to previous adverts. As a result, the marketer can determine whether different age groups respond differently within and between other age groups. It also enables the marketer to acknowledge the age group of participants as an extraneous factor which may obscure whether the wording actually made the difference. **Randomisation** may be implemented by randomly assigning one treatment to each participant within each age group. **Replication** is ensured by repeating the delivery of treatments multiple times within each age group. Once data has been collated, the marketer may conduct ANOVA calculations to either reject or accept the null hypothesis.

When we are working with randomised block experiments, such as that in Example 1.4, we may wish to construct an empirical model to quantitatively demonstrate the effect of treatments and blocking on the response (i.e., click-through rate in Example 1.4). Definition 1.5 describes a simple model for randomised block experiments.

Definition 1.5 (Linear Model for Randomised Block Experiments [MGM12]). Let an experiment comprise of T treatments, B blocks, and $T \times B$ units. Then, the response, Y_{ij}, of unit ij in block i undergoing treatment j, can be modelled by:

$$Y_{ij} = \mu + b_i + t_j + \epsilon_{ij}$$

where $\mu = \sum_{a,b} Y_{ab}/TB$ is the grand average of all the responses, b_i is the block effect of block i, t_j is the treatment effect of treatment j, and $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ is a random variable to account for the experimental error.

Notice that, in Definition 1.5, there does not exist a term to account for the influence of a person's friends and family on their response to a treatment. In the context of A/B testing, one would assume that there should be an element of social influence present [BSS23]. Therefore, accounting for such influence will lead to a more representative model of online A/B testing [PGS17].

1.3 Experimental Design on Networks

Parker et al. [PGS17] provide a model which adapts Definition 1.5 to networks, introducing a network effect to accommodate the social influence effect.

Definition 1.6 (Linear Network Effects Model [PGS17])**.** Let an experiment comprise of T treatments and N units. Then, the response, Y_i, of unit i undergoing treatment $t(i)$, can be modelled by:

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

where:

- $\mu = \sum_a Y_a/N$ is the grand average of all the responses
- $\tau_{t(i)}$ is the deviation of the response Y_i from μ as an inherent result of the treatment. This is known as the treatment effect of treatment $t(i)$ on unit i.
- A_{ik} is the ik-th element of the adjacency matrix (see Definition 1.2).
- $\gamma_{t(k)}$ is the influence from neighbouring nodes as a result of treatment assignment, known as the network effect of treatment $t(k)$ on unit k.
- $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ is a random variable to account for the experimental error.

The term involving the adjacency matrix (see Definition 1.2) allows for social influence from the nearest neighbours of each node. As such, the model described in Definition 1.6 can describe several social experiments in which participants may be influenced by those close to them.

The difficulties with implementing experiments on networks is that randomisation (see Remark 1.3(2)) relies on a fair random allocation of treatments. This becomes unreliable once one considers the topology of the graph (i.e., allocation of treatments may become biased if one treatment is allocated to nodes of relatively high degree). Furthermore, in a randomised experiment, we often require the same number of units allocated each treatment. This implies that a comprehensive structure of the network is required before solid experimental design takes place. However, we often do not know the full shape of the network as non-administrators, especially given the fast-changing diameter of large online networks like those platformed on Instagram or Twitter. Thus, more research needs to take place to further develop the theory of implementing online experiments on large networks.

2 Research Aims

- I wish to develop the theory around online experiments, with a particular focus on how one can conduct experiments on large networks. I recognise that there may be restricted randomisation tools we can use here to understand how to best allocate treatments.
- Develop theory around how to better implement online experiments on dynamic, fast-changing networks (see BA/preferential attachment model).
- Moreover, I wish to develop object-oriented Python code to illustrate how such an online experiment may be conducted, and its results may be analysed.

3 Research Plan

Month	Actions
1	• The first month was spent reading the first five chapters of Montgomery [Mon20] and the first four chapters of Mead et al. [MGM12] to understand the fundamentals of experimental design. This was essential given that I did not have not had any formal teaching on experimental design theory. • I also familiarised myself with Parker et al. [PGS17] to learn how experiments may be conducted on networks. • I reviewed the fundamentals of network science (adjacency, incidence, various random network models, preferential attachment model). • I have completed relevant asynchronous modules on Brightspace. • Upload weekly PhD diary entries and write posts on the basics of networks and experimental design.

Month	Actions
2	• I wish to study Chapter 11 of Mead et al. [MGM12] to be able to understand how to better assign random treatments in different ways. • I am aiming to develop my understanding of the wider context of online experiments, so will explore Larsen et al. review paper [Lar+24] in more detail. • Continue to develop my understanding of network structure, considering different ways of illustrating the macroscopic properties of a network. I wish to construct a glossary fo these different ways to see different approaches to fair treatment allocation.

Month	Actions
3	• Continue to develop my understanding of factorial experimental design, as this may lead to a different perspective on allocating treatments. • Check forward citations of Montgomery [Mon20] and Gilmour et al. to further understand the context of experimental design and its applications. • Develop understanding of R for data analysis. Use LinkedIn Learning for asynchronous materials. • Continue to write weekly PhD diary entries to demonstrate understanding and reflect on current progress.

Month	Actions
4	• Attend advanced Design of Experiments LTCC course: https://www.ltcc.ac.uk/media/london-taught-course-centre/Design-of-Experiments-Abstract.pdf • Attend module taught at Brunel to solidify understanding of ANOVA and R. • Develop understanding of R for network science. Great GitHub Pages site here: https://dshizuka.github.io/networkanalysis/index.html • Continue to write weekly PhD diary entries to demonstrate understanding and reflect on current progress.

Month	Actions
5	• Continue to attend advanced Design of Experiments LTCC course: https://www.ltcc.ac.uk/media/london-taught-course-centre/Design-of-Experiments-Abstract.pdf and work through problems to review understanding. • Further develop my understanding of blocking techniques and the impact that multiple blocking strategies can have on the accuracy of an experiment. • Upload tutorials on Python and R for network science and statistics. This will allow me to assess understanding and supplement a growing bank of notes. • Continue to attend module taught at Brunel to solidify understanding of ANOVA and R. • Begin to learn how factorial experiments can be analysed in practice. Perhaps write a post of the full process on my Github Pages site. • Continue to write weekly PhD diary entries to demonstrate understanding and reflect on current progress.

Month	Actions
6	• Reflect on understanding of LTCC course (https://www.ltcc.ac.uk/media/london-taught-course-centre/Design-of-Experiments-Abstract.pdf) and determine how the content of the module applies to experiments on networks. • Synthesise knowledge of blocking, randomisaton, and replication. Consider how this applies to networks. • Continue to attend module taught at Brunel to solidify understanding of ANOVA and R. • Develop understanding of R for network science. Great GitHub repository here: https://dshizuka.github.io/networkanalysis/index.html • Continue to write weekly PhD entries to demonstrate understanding and reflect on current progress.

Month	Actions
7	• Prepare to present at the Brunel Researcher Symposium to develop my dissemination skills. This will also be an ability to reflect on my current progress and consolidate what I have learned. • Investigate other opportunities to present such as those hosted by YSS. • Gain a better understanding on how to minimise variance of estimators within models. How does this apply to what I have studied so far about experiments on networks? Begin to evaluate different strategies and minimise variance to gain a better model. • Continue to write weekly PhD diary entries to demonstrate understanding and reflect on current progress.

Month	Actions
8	• Consider writing Python code which implements some of the strategies explored in the previous month • Consider how these scale up for larger networks. As in Parker et al. [PGS17], what happens when p is increased for Erdös Renyi networks? This is because I must consider not just the scale of the network (i.e., its diameter), but also how to implement strategies on faster moving networks. • Continue to write weekly PhD diary entries to demonstrate understanding and reflect on current progress.

References

[Bap10] B. P. Bapat. “Adjacency Matrix”. In: *Graphs and Matrices*. London: Springer London, 2010, pp. 25–44. ISBN: 978-1-84882-981-7. DOI: 10.1007/978-1-84882-981-7_3. URL: https://doi.org/10.1007/978-1-84882-981-7_3.

[BSS23] T. Bui, S. H. Steiner, and N. T. Stevens. “General Additive Network Effect Models”. In: *The New England Journal of Statistics in Data Science* 1.3 (2023), pp. 342–360. ISSN: 2693-7166. DOI: 10.51387/23-NEJSDS29.

[GLH14] J. Gao, D. Li, and S. Havlin. “From a single network to a network of networks”. In: *National Science Review* 1.3 (2014), pp. 346–356.

[Gos+18] M. Gosak et al. “Network science of biological systems at different scales: A review”. In: *Physics of life reviews* 24 (2018), pp. 118–135.

[Gra+15] P. A. Grabowicz et al. “Bayesian social influence in the online realm”. In: *arXiv preprint arXiv:1512.00770* (2015).

[KT17] R. Kohavi and S. Thomke. “The surprising power of online experiments”. In: *Harvard business review* 95.5 (2017), pp. 74–82.

[Lar+24] N. Larsen et al. “Statistical Challenges in Online Controlled Experiments: A Review of A/B Testing Methodology”. In: *The American Statistician* 78.2 (2024), pp. 135–149. DOI: 10.1080/00031305.2023.2257237.

[MGM12] R. Mead, S. Gilmour, and A. Mead. *Statistical Principles for the Design of Experiments: Applications to Real Experiments*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2012. ISBN: 9781139576642. URL: https://books.google.co.uk/books?id=Z_0LBAAAQBAJ.

[Mon20] D. Montgomery. *Design and Analysis of Experiments*. Wiley, 2020. ISBN: 9781119722106. URL: https://books.google.co.uk/books?id=kB7zDwAAQBAJ.

[PGS17] B. M. Parker, S. G. Gilmour, and J. Schormans. “Optimal design of experiments on connected units with application to social networks”. In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 66.3 (2017), pp. 455–480.

[Qui+24] F. Quin et al. "A/B testing: A systematic literature review". In: *Journal of Systems and Software* 211 (2024), p. 112011. ISSN: 0164-1212. DOI: https://doi.org/10.1016/j.jss.2024.112011.

[USK16] S. Urban, R. Sreenivasan, and V. Kannan. *It's all a/bout testing: The Netflix experimentation platform*. 2016. URL: https://netflixtechblog.com/its-all-a-bout-testing-the-netflix-experimentation-platform-4e1ca458c15.

[You14] S. W. H. Young. "Improving Library User Experience with A/B Testing: Principles and Process". In: *Weave: Journal of Library User Experience* 1.1 (Aug. 2014). ISSN: 2333-3316. DOI: 10.3998/weave.12535642.0001.101. URL: http://dx.doi.org/10.3998/weave.12535642.0001.101.