Book Marketing Strategy: Leveraging Networks

Li-Lian Ang

Minerva Schools at KGI

Table of Contents

Feedback and Grading

1. Was there a better way to design the simulation to achieve the results I was looking for?
2. Were the assumptions I made too outlandish? Was there a better way I could have justified them or a different distribution I might have used?

Book Marketing

Traditional advertising is both costly and ineffective for a self-published author without reputation or funds to boost the reach of their campaign. Social media provides a free or highly affordable alternative to buying ads in media like newspapers, radio commercials or billboards. The difficulty is getting visibility among the millions of posts from millions of other users and knowing how to reach your target audience in the digital jungle. This project will determine which marketing strategy best leverages social media networks to get the widest reach and sale to the target demographic, specifically to be applied to the historical fiction novel I will be releasing on 24 April 2021.

The proposed strategy is to host a giveaway where users can enter by sharing the book with their friends. At the end of the campaign, a set number of entries will be randomly selected to receive a free copy of my book. There are two methods of entering the giveaway which I aim to test: tagging friends on a post or sharing a post on their personal social media (i.e. through Instagram stories). The best method will provide the highest potential sales and reach to users who are currently unaware of the book. The former is determined by the persuasive ability of the entrant in getting their friends to perceive the book highly, and the latter by the number of new connections or followers to the book's social media accounts.

Additionally, I am looking to optimise several other parameters of the strategy, namely the number of people who should be tagged, the number of books to be given away and the length of the giveaway campaign. Even though maximising each of these parameters would intuitively give the best results, each comes with its constraints. Increasing the number of people to be tagged raises the 'cost' to enter the giveaway, which may dissuade users from participating.

Historical fiction is a relatively niche genre, so the most relevant users who should share the book are not necessarily the ones with the most connections. Similarly, the number of books for the giveaway is constrained by cost since the book's production and delivery costs can rise unbearably high depending on where the winner is located. Conversely, reducing the number of books too low might deter possible entrants if the odds seem too far out of their favour. Finally, a too-long campaign can drag out expectations and build negative emotions in entrants, while too short a campaign would fail to capture enough of an audience.

Rather than test out which strategy works best in reality where I will end up subject to the cost, effort and diminishing rewards from each iteration, I will attempt to derive a mathematical relationship between the input variables and test out my theory in a simulation.

Growth of Reach

Recurrence Relation

The principle of advertising is visibility. The more people who know about the product, the higher the likelihood they will buy it, even better if the people who know the product are part of the target audience. Therefore, with the book's marketing campaign, the reach will only grow. The question is by how much.

Tagging Friends

We can mathematically derive the expected growth in reach using a recurrence relation. If we assume that each follower has an equal probability of tagging their friends, we can define the total number of followers after a time step (i.e. one hour).

$$\begin{aligned} entrants &= P(\lambda) \\ &= P(p \cdot N) \end{aligned}$$

The first term for entrants defines the Poisson distribution probability governed by the rate parameter λ which in this scenario describes the expected number of entrants to the giveaway every hour. The distribution is appropriate to model the number of entrants since each event is independent of the other. It is often used to model discrete events that are randomly spaced. λ is based on the probability of each person joining the giveaway given by *p*.

$$followers\, per\, entrant \;=\; q \cdot \int_{0.5}^{1} \beta(1, 9)$$

The equation above defines the number of tagged people who will join the network. q denotes the number of people tagged the beta distribution models the intensity of support for the book for a tagged person, which shows their likelihood of entering into the giveaway themselves. The beta distribution is skewed to the left and assumes that most tagged users are not likely to engage, which is plausible given that the proportion of viewers translate to a small proportion of buyers. The assumption is conservative, giving the lower bound of how a tagged person's intensity of support would be.

Multiplying the two equations together will give the expected number of new followers. However, to simplify sampling from the Poisson distribution, we will use assume a uniform distribution and use $\lambda = pN$ to approximate the number of entrants. The following equation gives the total number of followers at each time step in a recurrence relation, where we assume each entrant only joins the giveaway once.

$$N_{n+1} = p \cdot N_n \cdot q \cdot \int_{0.5}^{1} \beta(1, 9) + (1 - p) \cdot N_n$$

We can simplify the recurrence relation:

$$N_{n+1} = p \cdot N_n \cdot q \cdot 0.00195 + (1 - p) \cdot N_n$$

$$N_{n+1} = N_n(0.00195pq + 1 - p)$$

Then, compute the common ratio r which gives us the rate of growth:

$$r = \frac{N_n(0.00195pq+1-p)}{N_n(0.00195pq+1-p)^2}$$

$$r = \frac{1}{0.00195pq+1-p}$$

Since p will always be between 0 and 1, the denominator will not exceed one except under extreme circumstances. Therefore, the common ratio will be more than one most of the time, showing a positive exponential growth in followers over time which is to be expected. The recurrence relation is said to be divergent, meaning that it will never stop increasing. The figure below shows that the number of friends tagged does not impact the follower growth rate except when we are more confident of high participation in the giveaway (i.e. when there is a high probability of tagging friends). It is interesting to note that at a high number of friends tagged, the rate of growth decreases which is unexpected since its increase is directly related to the growth rate.

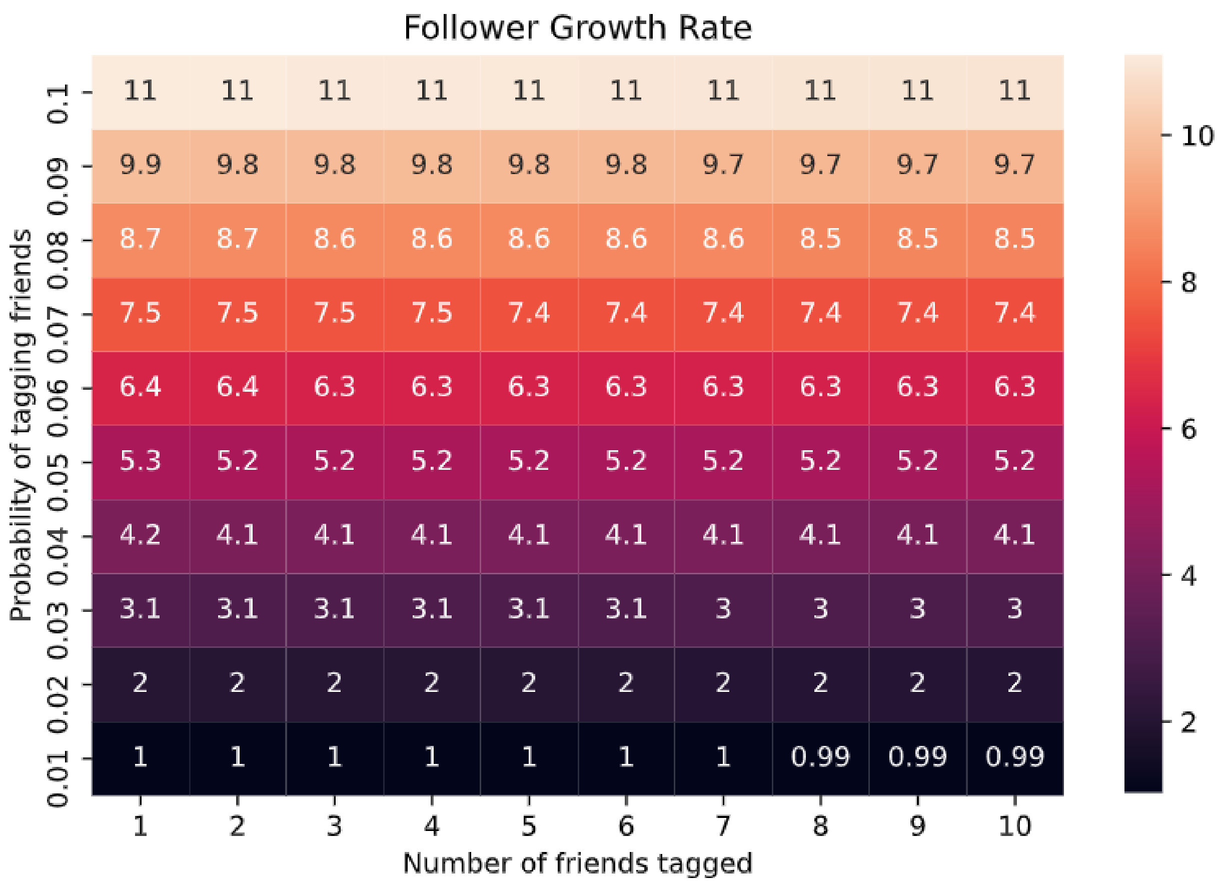


Figure 1. Heat map of follower growth rate in percentage as described by the common ratio equation for varying inputs of p (probability of tagging friends) and q (number of friends tagged).

These results are based on a flattening of the distributions we expect to see, which is appropriate for these calculations since we do not expect any emergent effects to occur given the simple mechanism. The implications for the results is to focus on increasing the desire for participation, possibly through increasing the number of books given away or including another desirable reward.

Posting Instagram Story

For this strategy, the recurrence relation is similar except that the people who successfully start to follow the book's Instagram are defined by:

$$followers\ per\ entrant\ = \int_{0.5}^{1} \beta(0.1, 9.9)\ \cdot n_{mean} \cdot q$$

q denotes the percentage of the posters followers who view their story. Similar to the previous strategy, the beta distribution models the intensity of support but has a far more extreme skew since the connection between viewers and the post are more tenuous with the less personal connection. The actual distribution of followers for each user is best modelled using a skewed normal distribution, but unfortunately it cannot be integrated, so we will use the average number of followers (n_{mean}) as a proxy.

The recurrence relation for the number of followers over time in this strategy is given by:

$$N_{n+1} = (p \cdot N_n \cdot \int_{0.5}^{1} \beta(0.1, 9.9)\ \cdot n_{mean} \cdot q)\ +\ (1\ -\ p)N_n$$

We can simplify the equation:

$$N_{n+1} = N_n(0.00002pqn_{mean} +\ 1\ -\ p)$$

Then compute the common ratio which shows the rate of follower growth:

$$r = \frac{N_n(0.00002pqn_{mean}+1-p)}{N_n(0.00002pqn_{mean}+1-p)^2}$$

$$r = \frac{1}{0.00002pqn_{mean}+1-p}$$

The relationship between the values is the same as before, except there is an extra term in the magnitude of 10^2, which makes the constant factor in the first term of the common ratio for this strategy ten times larger. This means that the expected reach for posting on Instagram stories is ten times more effective than tagging friends. These results seem plausible given that they work under the same assumptions with the beta distribution (i.e. probability of engagement) for this strategy much lower than the previous strategy.

Simulating Growth

To verify our mathematical model, we will simulate the social media network to determine if flattening the distributions created a significant difference in the expected results.

Instead of initialising the network with the number of nodes equal to the book's Instagram follower (250), we will use fifty as a proxy since a large initial value would require more computational power than is currently feasible for this exploration. Each node of these nodes will have an intensity of support for the book drawn from the distribution $\beta(2, 8)$, giving the intensity a left skew and constraining the values between 0 and 1. The distribution captures the relatively high degree of support from my initial followers, comprising my close contacts.

For the strategy of tagging friends, at each time step, all nodes will have a (p) probability of tagging their friends. Nodes who do decide to tag will tag *x* number of friends who, upon seeing the post, will each have support from the book given by the distribution $\beta(5, 5)$. To be realistic, we will experiment with values of p below 10%. If the tagged friend has an intensity of support over 0.5, it is assumed that they support the book enough to form a connection and potentially join the giveaway themselves. Therefore, they will connect with the node who tagged them with a weight given by the same beta distribution. The other nodes whose support falls below 0.5 are assumed to be uninterested.

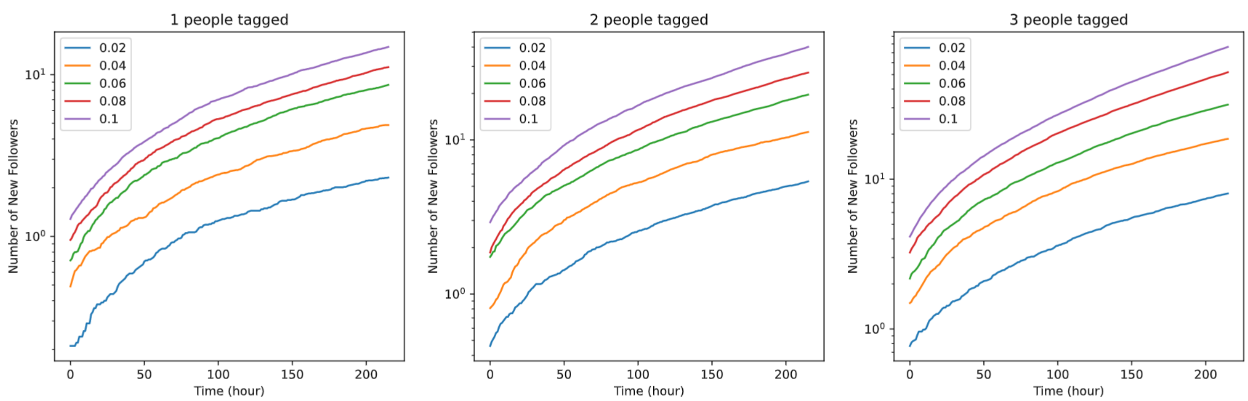
For the strategy of posting an Instagram story, in addition to intensity of support, each node will also be initialised with a follower count drawn from a skewed normal distribution with a mean of 300, based on the average number of followers of the book's Instagram followers, and a standard deviation of 200 which represents the high variability in follower count, ranging from 50 to almost 2000. The normal distribution is skewed by a factor of 4, following the perceived left-skewed distribution. However, since we scaled down our initial follower count by five, we will do the same here for consistency, so the mean will be 60 and the standard deviation 40.

Similar to the previous strategy, at each time step, every node has a probability of posting about the book on their story. The engagement rate for Instagram stories varies, but generally, it receives more views than regular posts. I will assume that 25% of each entrant's followers view the story, based on the percentage of views I usually get for stories on my personal Instagram account. Since each viewer only gives the story a cursory glance, the intensity of support for each viewer is given by the distribution $\beta(0.1, 9.9)$. Viewers who have an intensity higher than 0.5 will join the network.

The simulation also assumes that each user will only enter the giveaway once, so once they have tagged their friends or posted a story, they will not do so again. The assumption is conservative on how many times a person will enter their names into the giveaway.

The simulation aims to determine the best marketing strategy and the optimal campaign length, books to be given away, and the number of people to tag. According to Instagram, the book's Instagram account gets an almost uniform amount of interaction at any hour of the day, for any day of the week, which means that we can safely assume that the behaviour for each time step is equal to the average. The number of books to be given away can be thought to grow proportionally to the probability of a user entering the giveaway. Therefore, the parameters would give the greatest new followers.

Results



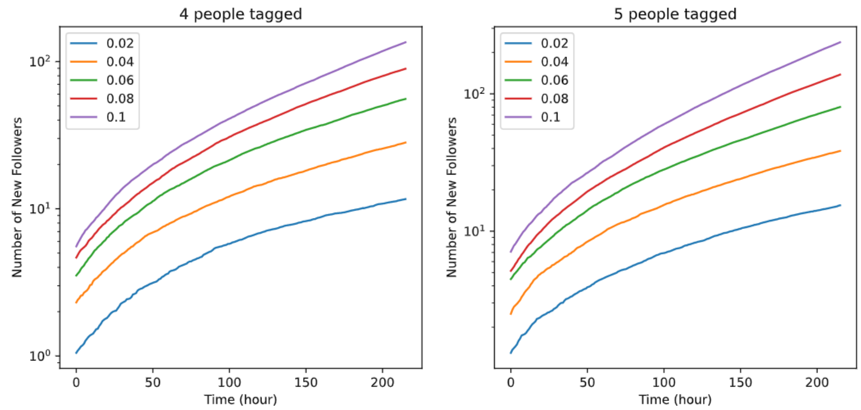


Figure 2. Growth of followers given varying people tagged per entry and probability of each person participating in a giveaway over a 10-day campaign period.

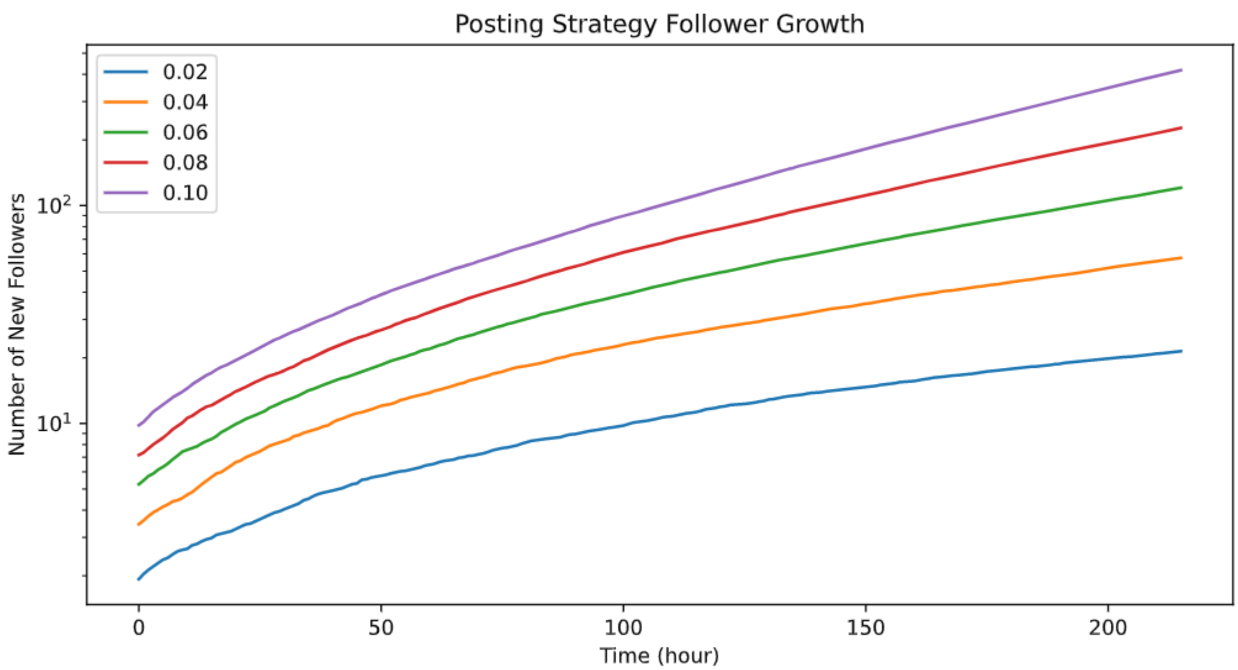


Figure 3. Growth of followers given varying probability of each person posting an Instagram story as part of the giveaway over a 10-day campaign period.

The simulation corroborates the positive correlation between the number of followers and the probability of participating in the giveaway. However, it does not reveal the geometric pattern from the mathematical model where we should expect to see a linear pattern since the y-axis is in log scale. This is likely caused by the added uncertainty from actually sampling the Poisson and beta distribution which compounded stifles the growth rate of new followers. Nevertheless, since we have used a much smaller number of the actual initial number of followers (50 instead of 300), we can expect the number of followers to grow to fairly high numbers.

Interestingly, the recurrence relation has a slight negative correlation between the number of people tagged and the growth rate of followers, which seems counterintuitive from the formula and real-life expectations. In the model, the number of people tagged significantly affects the total number of followers, more than doubling the number. This relationship might have been flattened out in the recurrence relation when assumptions were made on sampling for the beta and Poisson distribution.

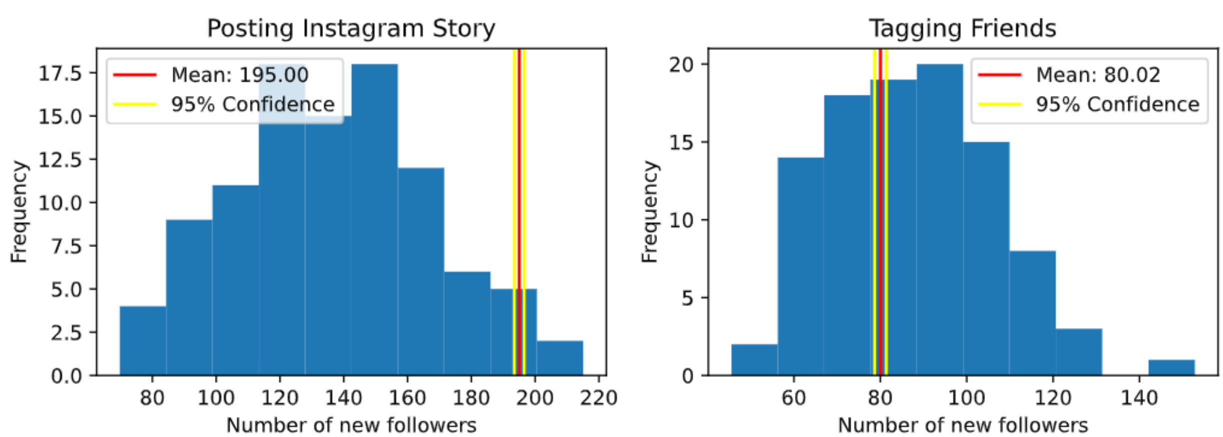


Figure 4. Distribution of number of new followers over 100 trials of simulation for probability of participating in giveaway at 1% with 95% confidence intervals. The number of tagged friends for the figure on the right is 5.

The confidence intervals for the means of both strategies is fairly narrow showing high certainty in our values as the true value. However, it should be noted that especially for the tagging friends strategy, the mean is extremely sensitive to the beta distribution used to approximate the intensity of support of new nodes. A shift from $\beta(1, 9)$ to $\beta(2, 8)$ yielded an increase to the magnitude of 10^1. The same is true for the posting Instagram story strategy.

Increasing Sales

The best marketing strategy will give the greatest sale of books which can be approximated based on the number of nodes with an intensity above 0.8 where we can assume that they support the book enough to buy it. If there is a marginal difference between the number of nodes, we will look at the distribution of intensity for the distribution with the most left skew.

Simulating Sales

This simulation builds on the previous one, where each pair of nodes was initialised with connections to each other following the preferential attachment defined by Barabasi-Albert. In this setup, nodes with a higher initial degree have a higher likelihood of forming new connections, which follows that popular people tend to make connections more quickly than less popular people. The connection also captures the observation that most of my followers are acquainted with each other by being a part of the same social circles. The edge between each node is weighted using the distribution $\beta(6, 4)$, which gives a less intense right skew than the previous distribution. The weights represent the strength of connections between nodes and, by extension, how much nodes can influence each other's support for the book. The distribution reflects my more conservative assumption of the relationships between my followers than my relationship with my followers.

Additionally, every new node which joins the network is randomly connected to one other node in the network by preferential attachment to simulate the small world network where friends tend to have several mutual connections. The weighted edge between an existing node and a new node is given by $\beta(5, 5)$, which reflects a conservative expectation of how much influence friends have over each other. The weighted edge between the new node and a random node in the network is given by a uniform distribution that is more even to reflect the uncertainty of various connections between nodes.

Given that the campaign will cause interactions between nodes (between an entrant and the friends they tag or between a poster and their followers who respond to their story), the relationship between each of these nodes and the book will change as well. Positive sentiments from one node to another could increase their support for the book, for example, if they've already read the book and recommend it to their friends. Whereas negative sentiments could decrease and even break off connection to the book, if, for instance, they had a bad experience with the book.[1] An individual sentiment could spill over to their neighbours dependent on the strength of the relationship between two nodes, denoted by β, and how susceptible their opinions are to change denoted by α.

The change in opinion of one node interacting with another is given by $\Delta\sigma_i = \alpha w_{ij}(\sigma_j - \sigma_i)$, while the change in the relationship between two nodes after the interaction is given by $\Delta w_{ij} = \beta w_{ij}(1 - w_{ij})(1 - \gamma|\sigma_i - \sigma_j|)$. γ controls how much their opinion difference matters.

In this simulation, α is relatively higher than the rest of the parameters since people tend to take more stock in reviews given by friends or people they know. β has a low value since the interaction between nodes is minimal. Over such a short time, it is unlikely to affect their relationship. γ is given a value lesser than 1 to dampen the effect of the opinion difference given the low interaction between nodes.

[1] The implemented simulation will not be simulating breaks in connections between nodes.

After the tenth day from the simulation on the growth of reach, we will update the relationships between nodes over 30 days once the network stops growing. The unit for the timestep has increased since it is unlikely for relationships and opinions to change within the space of hours. It is more realistic to expect a change over days.

Result

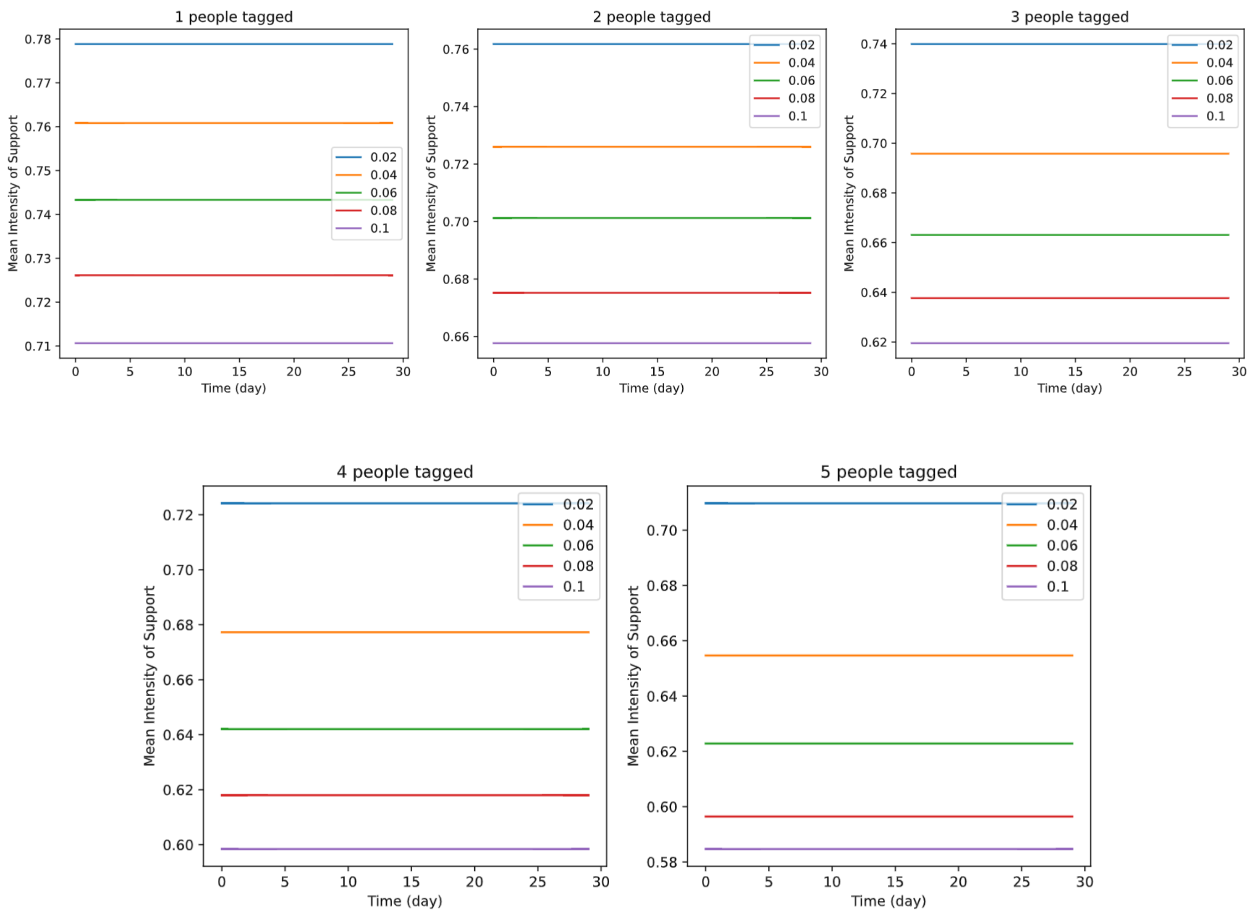


Figure 5. Mean of the intensity of support across all followers over 30 days for the varying probability of participating and number of people tagged.

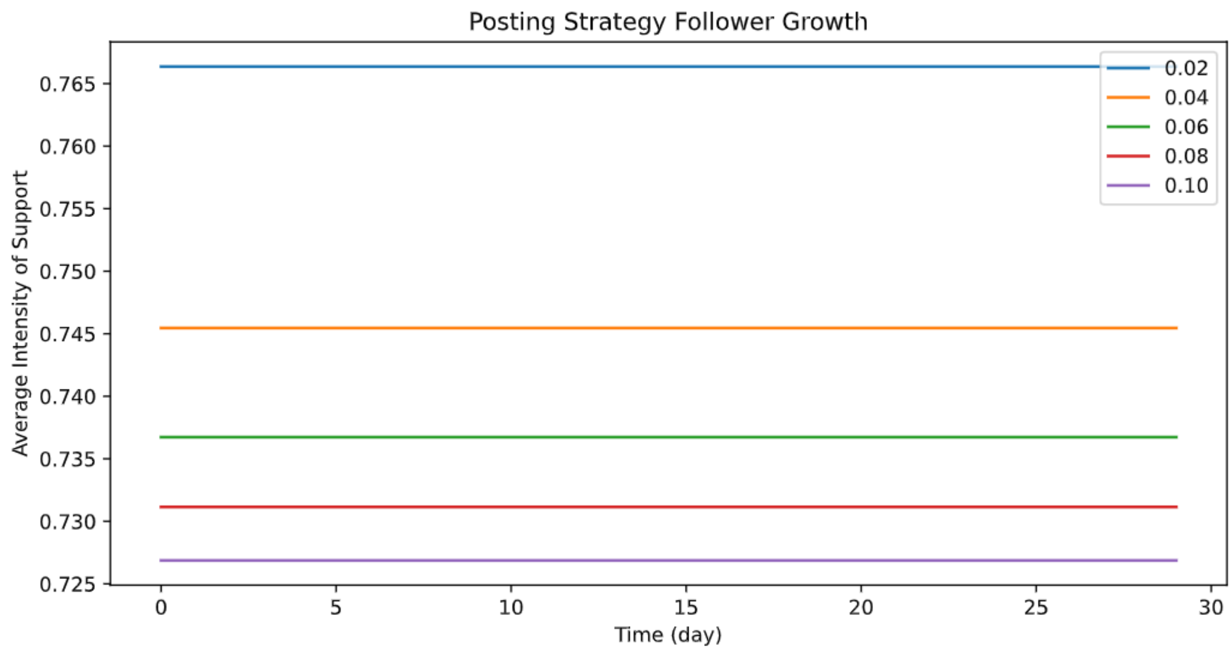


Figure 6. Mean of the intensity of support across all followers over 30 days for varying probability of participating.

The figures above show that the mean intensity of support does not shift discernibly, but it does become more certain as the standard deviation decreases over time across all variations of the probability of participating and the number of people tagged for both strategies. Interestingly, the results reveal that as the probability of participating increases, the overall average intensity of support decreases, likely because of the influence of new nodes entering the network, which dampen the intensity of support of existing nodes. This makes sense since each new node that enters the network will, on average, have a lower level of support for the book, so when there are multiple nodes with lower support, it will overwhelm the initially supportive node, thus bringing the average of the network down. We can see this pattern as the horizontal lines become slightly lower as the number of people tagged increases. Overall, it appears that the nodes exert an equalising force on each other than stagnate the average intensity of support rather than causing it to fluctuate.

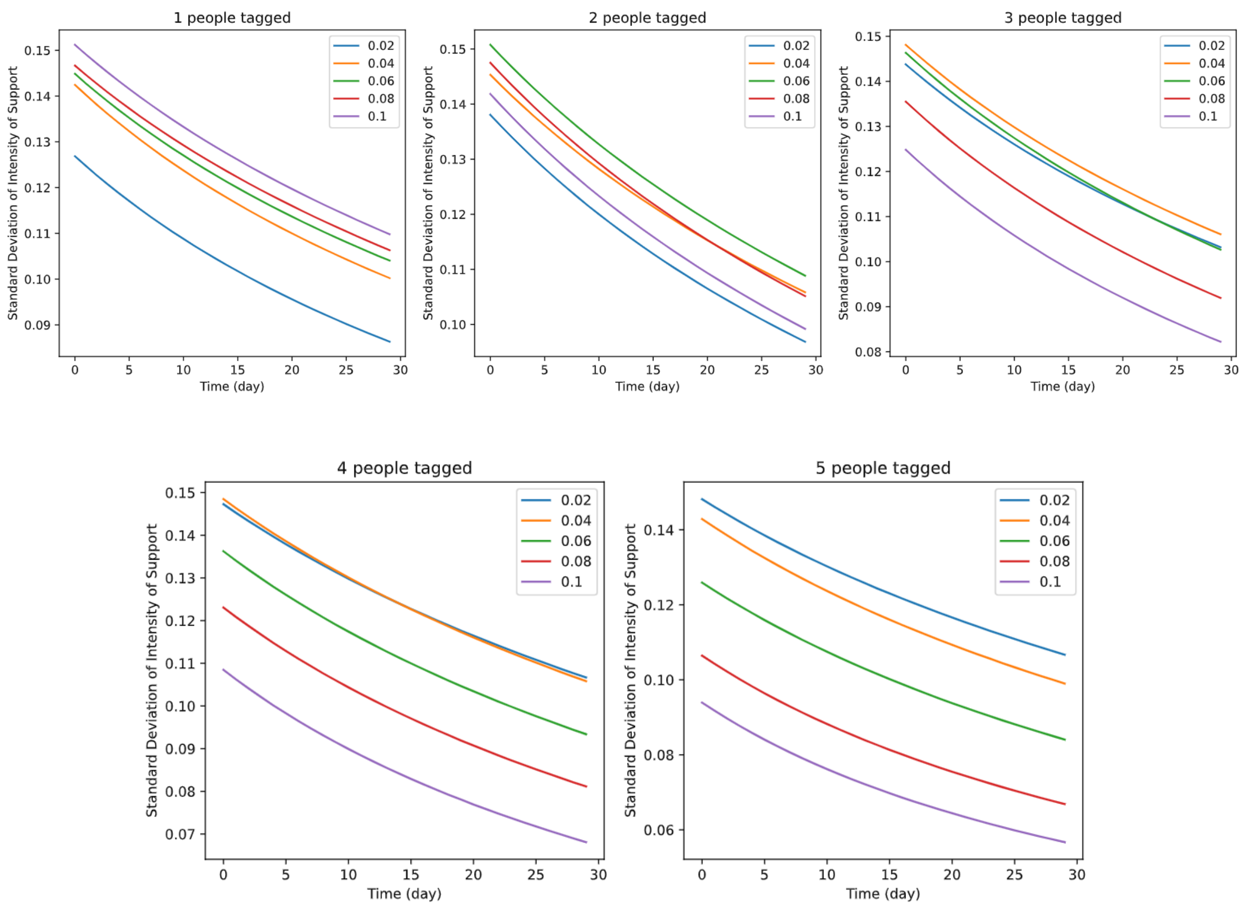


Figure 7. The standard deviation of intensity of support across all followers over 30 days for varying probability of participating and number of people tagged.

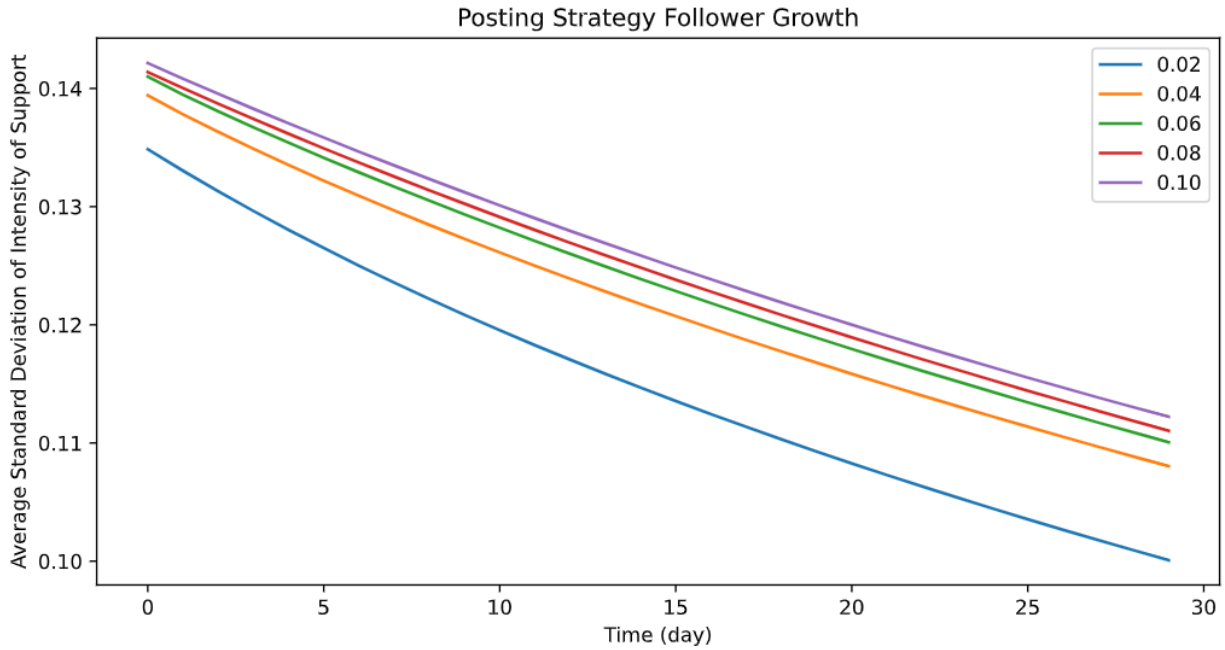


Figure 8. The standard deviation of intensity of support across all followers over 30 days for varying probability of participating.

The standard deviation across the number of people tagged and probability of participating for both strategies decreases at equivalent rates which shows that there isn't a significant skew or spread in the distribution of intensity of support. We can assume that the values are very close to the given mean.

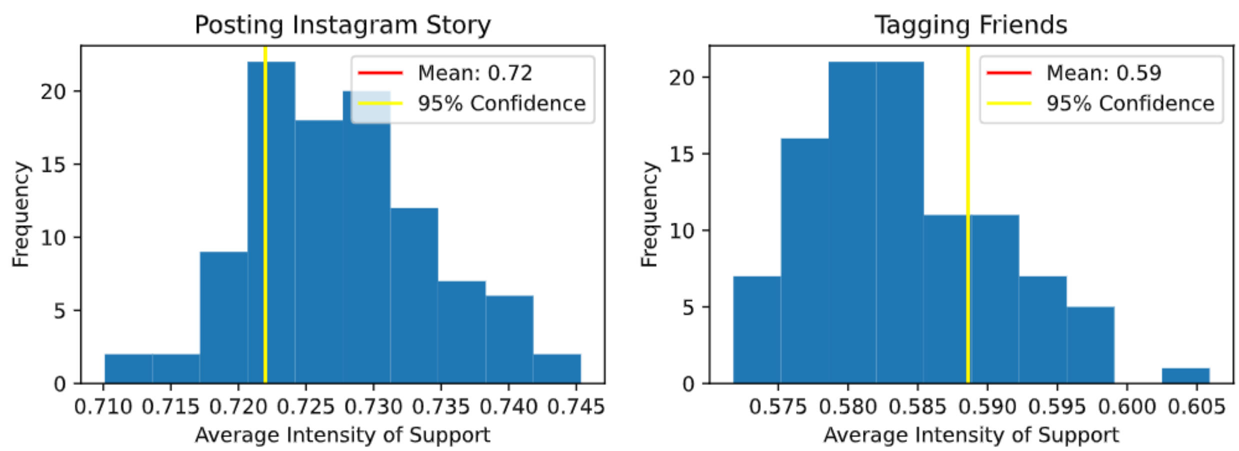


Figure 9. Distribution of average intensity of support for the probability of participating at 1% with 95% confidence interval. The number of people tagged for the figure on the right is 5.

There is a narrow confidence interval for the average intensity of support which shows that the results are close to the true value. However, the simulation for the change in intensity of support does not appear to appropriately model support for a product in the same way it might model the change of ideas. There usually aren't very strong feelings associated with them or much space to debate them which could lead to worsened relationships. Instead, it shows how apathy can be spread to the main supporters of the book as the average intensity of support decreases across the board. I believe that the assumption underlying the change in support to the book is fundamentally erroneous and would disregard them.

Taggers vs Posters

Metric	*Tagger (Tag 5)*	*Poster*
Number of new followers	80.02	195.00
Intensity of Support	0.59	0.73

Table 1. Summary of output variables for the probability of participating in the giveaway at 1% for tagging and posting strategy.

With these results, it would appear that posting an Instagram story is the best marketing strategy to gain the widest reach and the sales as approximated by the metric for the number of new followers and intensity of support, respectively. However, these results are extremely sensitive to the assumption of the distribution of intensity of support for new nodes, so I would not take these results at face value in deciding which strategy to ultimately employ.

There are two key insights to be gained. Firstly, the growth in reach increases exponentially with time, so it is worth allowing the campaign to run for a longer period or at least for 4 days when the growth in reach begins slowing. Secondly, the simulation shows that increasing the number of people tagged provides minor gains to the reach. If we factor in the additional effort it will take to tag more people with the probability of a person participating in the giveaway, the effect might cancel each other out. So it is better to err on the safe side in the number of people to tag.

In conclusion, the decision for the marketing strategy depends on my prior belief in how much more likely tagging friends or posting an Instagram story is to attract a higher number of interested and relevant demographic for the book. While posting an Instagram story may make the post visible to more people, it is not as engaging as being tagged in a post. The former would leave the selection process to chance while the latter gives my followers the opportunity to select interested parties, drawing on the wisdom of the crowd to pull the right demography to my book. This exploration has given helpful insights in creating content knowing that the success of the campaign can be significantly improved by eliciting a high intensity of support from the first impression possibly through offering more books in the giveaway. The cost may be worth the reward.