An Agent Based Model of the Majority Illusion

Jake Angell[1], Michael Cataldo[2], and Matthew Chamberlain[3]

[1] School of Computing, Teesside University
m2012929@live.tees.ac.uk
[2] School of Computing, Teesside University
g7130816@live.tees.ac.uk
[3] School of Computing, Teesside University
m2014976@live.tees.ac.uk

Abstract. This paper explores the hypothesis that the majority illusion paradox is stronger in networks where active nodes have a higher degree. As a result of the paradox, the spread of a real world social bias may be overrepresented within a community. The effect of degree attribute correlation on the strength of the paradox in synthetic networks is explored using an agent based approach. We demonstrate empirically that in most cases the majority illusion is stronger in networks where high degree nodes are more likely to be active. Activating high degree nodes influences the local observations of many nodes, resulting in a stronger majority illusion.

1 Background

While the paradoxical nature of social networks has been a significant area of research for some time, Lerman et al. (2015) have recently discovered a related paradox which they refer to as the “Majority Illusion”. This is the phenomenon where local behaviour is perceived to be more widespread globally than it actually is. People assume that views or behaviours held in their social group are also held by a wider majority and, as a result, adopt the views or behaviours themselves. In some cases, the paradox may be influenced by the underlying structure of the social network which can often skew the observations of individuals. Therefore, social behaviours that exist among a minority can often spread throughout a network.

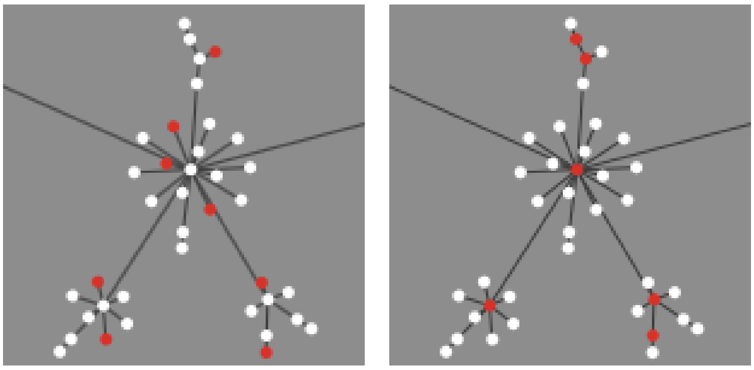
The majority illusion can be seen in real life networks, even more noticeably on websites such as Facebook and Twitter. The majority illusion effectively makes an opinion or behaviour seem more or less popular than it actually is. The research discussed in this paper explores how initial network configurations impact the spread of some behavioural trait, with a focus on degree-attribute correlation. Gaining a deep understanding of the ideal configuration for optimal behaviour spread will be lucrative for many, including public relations companies and even individuals with a large online presence.

Aside from the actual algorithm used to create a network, additional properties work conjointly to alter an existing network structure. One such property is assortativity, which is a measure of a node’s preference to connect to either similar or dissimilar nodes. Degree attribute correlation, is a property that does not directly affect the network structure, but in this paper we will certainly demonstrate that in some cases it has an effect on the spread of social contagions. Degree attribute correlation is the relationship between a node’s degree and how likely they are to be active. For example, the hubs within a network with a high degree-attribute correlation would be more likely to be active.

2 Introduction

As an alternative approach to the theoretical analysis of the Majority Illusion as presented by Lerman et al. (2015), we simulate the paradox in an agent based environment. This paper aims to build upon the mathematical model by evaluating the spread of an attribute over a network, similar to Granovetter (1978). The state of the model is determined by the result of an agent population making decisions at each time step in the model. An agent’s state will affect the subsequent decisions of its neighbours which may lead to propagation of attributes through a network. Agent state is defined as either being active or inactive and a threshold model is used to represent the amount of social proof an agent requires before switching to an active state. This type of agent deliberation may result in an emergence of behaviour which is not explicitly defined or expected.

In accordance with Lerman et al. (2015), we use a constant activation threshold ϕ, of 0.5, so that agents need a majority of their connections to be active to adopt the active state themselves. The model uses Erdős-Rényi and scale-free networks in a controlled environment of 2000 agents to study how network structure influences the strength of the Majority Illusion paradox. The two network structures are used as a foundation to explore how degree attribute correlation, referred to in this paper as $k - x$ correlation, influences agent behaviour and by extension, the adoption of active state. Using Erdős-Rényi, we generated a random graph of connected agents in order to determine the strength of the paradox in a non-assortative network. The findings are then compared with the results of disassortative, scale-free graphs. Furthermore, each graph type is tested in isolation with the use of additional, lower level network properties such as $k - x$ correlation.



(a) Low $k - x$ Correlation (b) High $k - x$ Correlation

Figure 1a, shows a small scale-free network with low $k - x$ correlation where the active nodes have very little influence due to their low degree. Inversely (1b) shows a high $k - x$ correlation which distributes the active attributes to the higher degree nodes. It can be observed that the majority illusion is significantly stronger in (1b).

3 Model

The model is based on an agent population that are connected to form various network structures. In this model, active agents are coloured red and the connections between agents are represented by lines. Agent deliberation takes place which is governed by the state and number of their connections. The model has been developed using NetLogo (Figure 2) and provides the following functionality to create various network structures with different properties.

3.1 Setup

The model presents multiple setup functions for the purposes of testing various network configurations. This includes the ability to select different network types, each using a corresponding algorithm for generating the structure of the network.

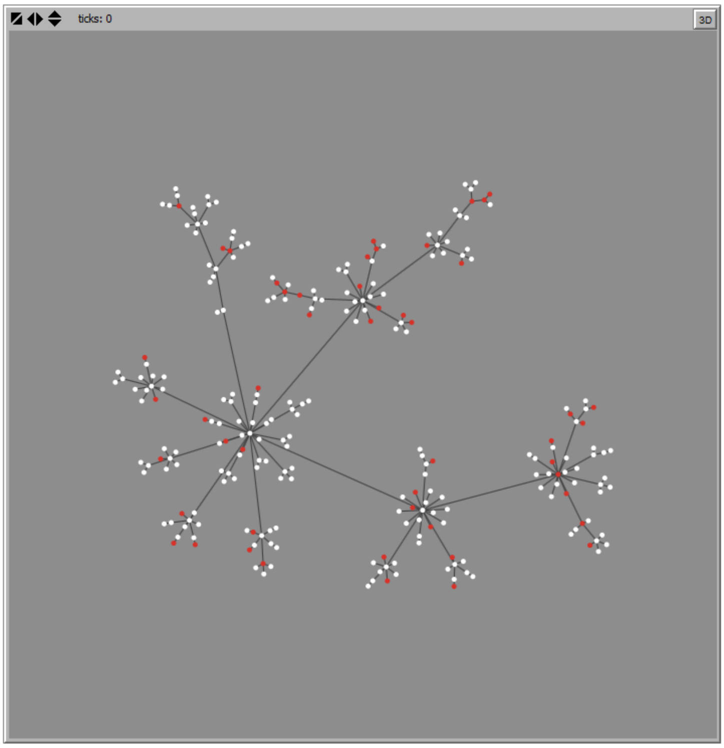


Fig. 2: A randomly generated Scale-free network using NetLogo

Erdős-Rényi networks are generated using a random Poisson degree distribution derived from a specified network wide degree average, ⟨k⟩.

```
Assign each agent a degree value from a Poisson distribution
foreach agents in network do
    while node connections < agent degree value do
        attach to another agent whose connections are less than their degree value
    end
end
```

Algorithm 1: Random Graph with Degree Distribution

Scale-free networks are generated using a preferential attachment mechanism as outlined in the Barabasi-Albert model (Albert & Barabasi 2002). The nature of this process inherently gives highly connected nodes more chance of receiving new connections. Algorithm 2 shows this process:

```
Start with a network with two connected nodes
foreach agent in agentset do
    Randomly select an existing network edge
    Randomly select a node on one end of edge
    Create a connection between node and the random node
end
```

Algorithm 2: Preferential Attachment Mechanism

Once an initial network structure is set up the model can apply several iterative processes to alter the network to satisfy a desired network property.

3.2 Setup with assortativity

Setup with assortativity uses the base setup function and then applies degree preserving edge-rewiring (Xulvi-Brunet & Sokolov 2005) until a desired assortativity value is achieved. To compute an assortativity value r for the network, the following equation is applied as given by (Newman 2002):

$$r = \frac{1}{\sigma_q^2} \sum_{jk} jk \left(e_{jk} - q_j q_k\right)$$

This can be rewritten in the following form for the purpose of implementing a practical model. M is the total number of edges. J and K are the remaining degree of the vertices at each end of a selected edge:

$$r = \frac{M^{-1}\sum_i j_i k_i - \left[M^{-1}\sum_i \frac{1}{2}(j_i+k_i)\right]^2}{M^{-1}\sum_i \frac{1}{2}\left(j_i^2+k_i^2\right) - \left[M^{-1}\sum_i \frac{1}{2}(j_i+k_i)\right]^2}$$

Algorithm 3 shows the edge rewiring procedure that is used in the model which uses Newman's formula at each iteration.

```
Start with the existing network
while current network r ≠ desired r do
    Randomly select two network edges and remove them
    Sort the four agents on the ends of each edge by degree in ascending order
    if current network r < desired r then
        connect the most like degree agents
    else
        connect the least like degree agents
    end
end
```

Algorithm 3: Edge-Rewiring Procedure

3.3 Setup with degree distribution

Setup with degree distribution uses the base setup function and then applies an attribute swapping procedure to achieve a desired degree attribute correlation. Randomly activating nodes creates a configuration with a $k - x$ correlation close to zero. This procedure preserves the percent of activeness across the network as shown in Algorithm 4.

```
Start with initial network
while current network k − x ≠ desired k − x do
    Select a random active agent (A)
    Select a random inactive agent (B)
    if desired k − x > current network k − x then
        if degree of A < degree of B then
            swap attribute values
        end
    else
        if degree of A > degree of B then
            swap attribute values
        end
    end
end
```

Algorithm 4: Attribute Swapping Procedure

3.4 Setup with assortativity and degree distribution

Setup with degree distribution and assortativity produces a network and then applies both degree preserving edge rewiring (See Algorithm 3) and attribute swapping (See Algorithm 4) to achieve both the desired assortativity and degree attribute correlation value.

3.5 Network Properties

The network has functionality to control several network properties:

- **Total Nodes**
  The total number of nodes used within the network. This can range between zero and ten thousand.

- **Average Degree**
  The average degree of each node throughout the network. The degree average can range between the values one and ten and is only used within Erdős-Rényi networks.

- **Degree Attribute Correlation**
  This property represents the relationship between the distribution of active nodes and the degree of those nodes. This property varies between zero and one, as it gets closer to one nodes with higher degrees are initially much more likely to be active.

- **Assortativity**
  A value indicating how the nodes of the network are connected considering their degree. The assortativity ranges from -1 to 1 where 1 means nodes of a like-degree are more likely to connect and -1 means nodes of unlike-degree are more likely to connect. As we are using degree preserving edge rewiring it is often difficult to achieve an exact value for assortativity given certain network structures.

- **Threshold Value**
  A value indicating the amount of social proof a node needs to change its attribute to active. As the percentage value lowers the agent is more easily influenced. A 50% threshold means that the majority of a node's neighbours need to be active in order for that node to be influenced by the paradox.

- **Group Decision Tick**
  This boolean value is used to determine the way that the model functions on a per tick basis. If the value is false, then each tick one agent makes a decision. If the value if true, then every agent makes a decision each tick.

4 Results

Through using NetLogo's Behaviour Space, we obtained test data which allowed systematic experiments to be carried out on the model using a given test definition. Before each test, several control variables had to be satisfied. These include assortativity, average degree and $k-x$ correlation. We used this approach to test the same initial network structure with differing $k-x$ values. Each test case was repeated and the mean of the results were taken to ensure more reliable results for analysis.

We quantified the strength of the majority illusion as the percentage of active nodes at model completion. Several controlled network structures and a varying $k-x$ correlation were used to explore the strength of the majority illusion.

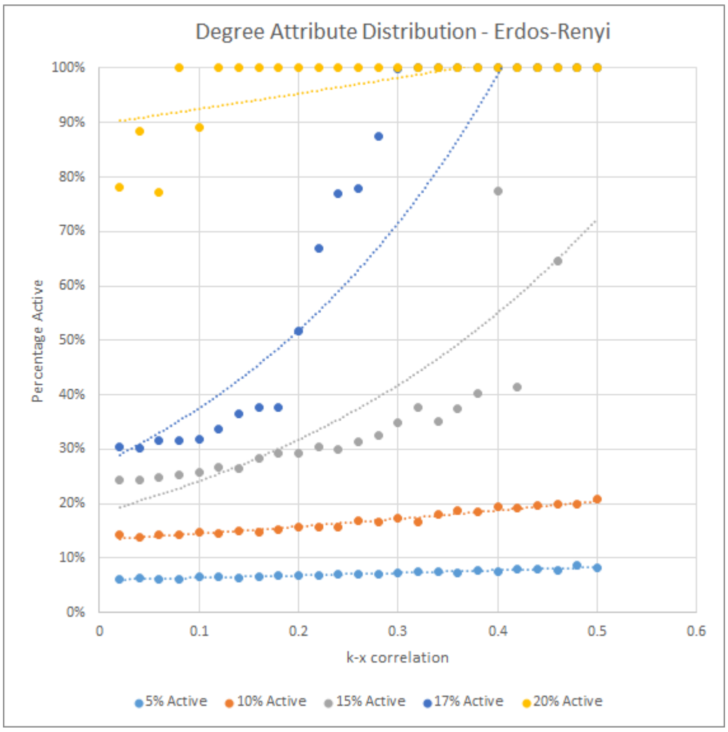


Fig. 3: presents the percentage of active agents at the conclusion of a run as a function of the $k - x$ correlation for different values of $P(x = 1)$. Each network has 2000 nodes, with $\langle \mathrm{k} \rangle = 4$. Each point represents a concluded run of the model.

At low values for the percentage of initially active nodes, increasing the $k-x$ correlation has little to no impact on the strength of the majority illusion. However, a tipping point where a significant change in behaviour is identified when $k-x$ correlation is 0.18 and the initial active percentage is 17%.

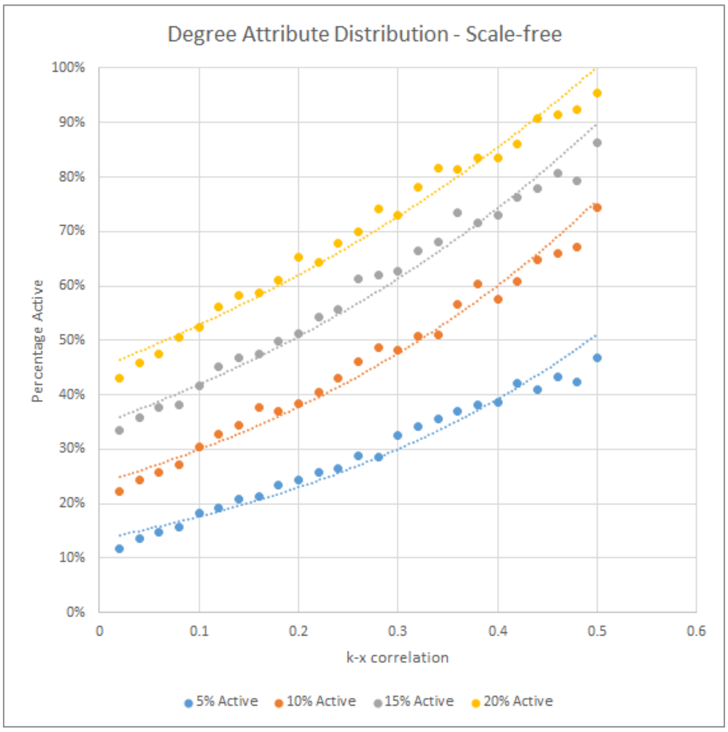


Fig. 4: presents the percentage of active agents at the conclusion of a run as a function of the $k-x$ correlation for different values of k. Each network has 2000 nodes, with $P(x=1)=0.05$, and $r=0$. Each point represents a concluded run of the model.

It can be taken from Figure 4 that increasing the $k-x$ correlation in a scale-free network such as this one should have a somewhat linear, positive effect on the eventual percentage of active nodes within the network. Altering the percentage of initially active nodes does not seem to amplify the effect of $k-x$ correlation in any way. In other words, if a general conclusion was to be drawn from Figure 4 it would be that increasing $k-x$ correlation in scale-free networks increases the final percentage of active nodes, regardless of the percentage of initially active nodes.

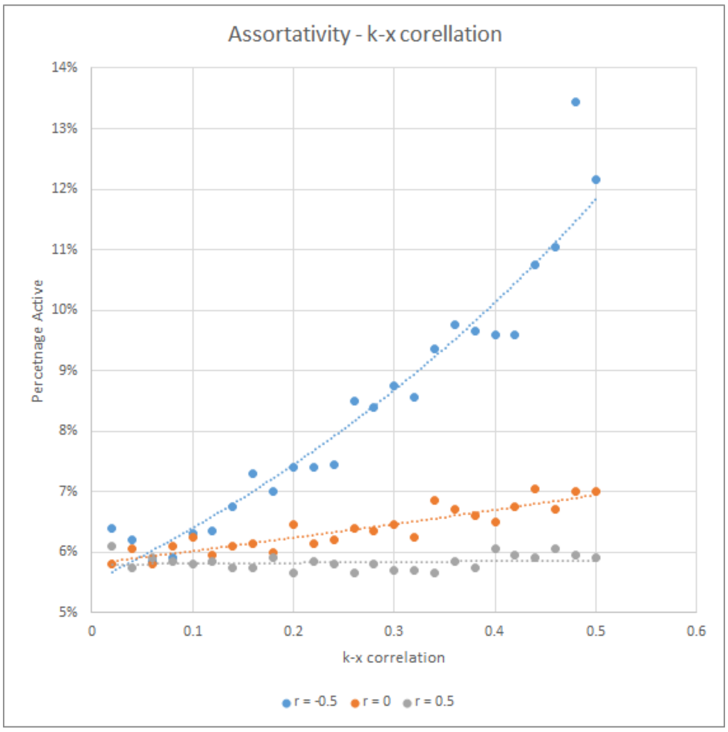


Fig. 5: plots the percentage of active agents at the conclusion of a run as a function of the $k - x$ correlation with different values for percentage of initially active nodes. Each network has 2000 nodes with ⟨k⟩= 4 and $r = 0$. Each point represents a concluded run of the model.

Figure 5 shows that the disassortative network is the most responsive to an increasing $k - x$ correlation, with the non-assortative network showing a slight increase and the assortative network showing very little or no response. From this it is reasonable to conclude that increasing the $k - x$ correlation in disassortative networks increases the strength of the majority illusion.

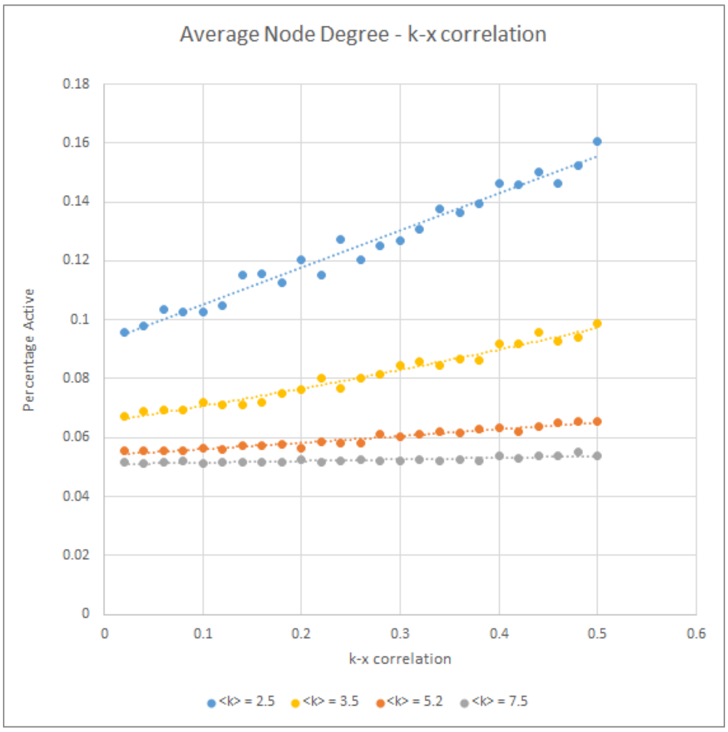


Fig. 6: plots the percentage of active agents at the conclusion of a run as a function of the $k-x$ correlation with different values for assortativity. Each network has 2000 nodes with $\langle \mathrm{k} \rangle = 4$ and $P(x=1) = 0.05$. Each point represents a concluded run of the model.

As the average degree of the network increases it can be concluded that $k-x$ correlation has less of an effect on the strength of the majority illusion. In other words, increasing the $k-x$ correlation in low average degree networks will exacerbate the majority illusion. Little to no effect on the strength of the majority illusion is observed when the $k-x$ correlation is changed within networks where the average degree is greater than 6.

5 Evaluation

Lerman et al. (2015) states that the majority illusion can be ultimately traced to the power of high degree nodes having the ability to skew the observations of many other nodes. Altering the $k-x$ correlation provides a way to control how active attributes are distributed among the nodes based on their degree. Therefore, monitoring the effects of different values for $k-x$ correlation forms a foundation to test this fundamental aspect of the majority illusion. In other words, the tests included in this paper are focused on testing the hypothesis that increasing the number of high degree, active nodes exacerbates the majority illusion.

The results collected generally conform to the hypothesis, however there are some specific cases where increasing the $k-x$ correlation seems to have little to no effect. This type of behaviour or lack thereof is usually seen when the network structure in question has some other, restricting property. Although, it is fair to expect that the $k-x$ correlation will not have an effect on some networks with these extreme properties. For an exaggerated example, consider a network where every node has a degree of two, it is evident that changing the $k-x$ correlation in this network would result in no change. Despite this observation, when tested against $k-x$ correlation, every network property lead to a positive effect on the strength of the majority illusion.

In networks with positive assortativity, nodes with similar degree are more likely to be connected to each other. Thus high degree nodes are more likely to connect to other high degree nodes. As high degree nodes are the least impressionable, a singular connected node will not have much influence over them during deliberation. Therefore the effects of increasing $k - x$ correlation is inconsequential, as observed in Figure 5.

The results show that the majority illusion and therefore the spread of a social contagion is generally stronger in scale-free networks compared to Erdős-Rényi networks. This is to be expected as scale-free networks are naturally disassortative so it is logical that increasing the amount of hubs that are active will have a positive impact on the majority illusion. However, Erdős-Rényi networks start to perform better due to the tipping point achieved by using high values of $k - x$ correlation in conjunction with a large percentage of initially active nodes. This finding is not trivial and requires comparatively more thought than one would expect. One possible explanation for this finding is that it is easier for a cumulative effect to thrive within an Erdős-Rényi network with a high average degree compared to its scale-free counterpart. As scale free networks are disassortative, once a low degree node becomes active it is unlikely that this change would have an effect on subsequent nodes. Erdős-Rényi networks are non-assortative and degree distribution can be specified more easily, thus it can be assumed that when a node becomes active it will have a greater effect on the observations of other, neighbouring nodes. This explanation is fully theoretical so it may be used as a basis for future research.

5.1 Limitations

Due to the fact that assortativity is achieved via degree preserving edge rewiring it becomes extremely hard to attain certain specified values. This process is also time consuming so it is not feasible to run many tests when the desired assortativity value is far from the value generated by the structure before edge rewiring. As assortativity is not a main concern in this paper, the edge rewiring procedure is sufficient. If assortativity was to be examined in more detail then optimisations would be made, most probably by considering assortativity as the network is constructed.

We assume that every agent is equally impressionable, whereas in a real life scenario, this is unlikely to be the case. Granovetter (1978) examines variance of threshold ϕ amongst a population, emulating behaviour present in social networks. Incorporating this may help to substantiate further investigation using an agent based approach.

The performance of the model becomes an issue when the node count is increased. Although, this is not a huge concern as all of the variables used in the model work on a percentage basis so there should not be large discrepancies between runs with distinct node counts.

6 Critique & Conclusion

The results gathered point to the conclusion that the majority illusion and therefore spread of a social contagion is stronger in networks where the highly connected nodes tend to be active, unless another limiting property is present. The results presented in this paper mostly show clear, well defined curves where the only differing network property is $k - x$ correlation. The property of $k - x$ correlation is also self-limiting, in that altering it does not have an affect on any other property in the network, so it is fair to conclude that it is solely responsible for the findings presented. Although, variations may be seen in behaviour over multiple runs of the model with the same network properties. This issue was addressed by performing multiple runs for each test and averaging the results. Overall, the findings of this paper align with the general sentiment in the paper of Lerman et al. (2015), although additional dynamics have been uncovered here that focus more on the spread of the contagion over time. Continued research on the intricate workings of the phenomena surrounding network configurations and the spread of behaviour is of vital importance, especially with the continued growth of online social networks.

References

Albert, R. & Barabasi, A. (2002), 'Statistical mechanics of complex networks', *Reviews of Modern Physics* .

Granovetter, M. (1978), 'Threshold models of collective behaviour', **83**.

Lerman, K., Yan, X. & Wu, X. (2015), 'The majority illusion in social networks', *USC Information Sciences Institute* .

Newman, M. E. J. (2002), 'Assortative mixing in networks'.

Xulvi-Brunet, R. & Sokolov, I. M. (2005), 'Changing correlations in networks: Assortativity and dissassortativity'.