Reducing the Propagation of Misinformation in Social Networks with Agent-Based Modelling Methods

Kyle Russell, 13831056
AUT University

June 20, 2017

Abstract

Misinformation is a historic social problem that has the potential of provoking large emotional responses which are often misguided. With the recent emergence and widespread participation of online social-networks, the dissemination of misinformation has been accelerated by exploiting our exposure to information and the trust we place in our social structures. The behaviour and characteristics of misinformation are well-researched however much is still unknown about the methods in which the propagation of misinformation in social-networks can be reduced. In this paper we investigate the problem using an agent-based approach where we develop an agent-based model to capture the characteristics of online social-networks with genuine and fake information circulating, where we model agents based on sharing online users. Agents were given decision making mechanisms where they were able to optimally choose which neighbouring agent to influence as well as deciding which agent to be influenced by. Agents autonomously propagate their information to neighbouring agents in a network who are influenced and adopt the information of their influencer where they are not aware of the reliability of the information they adopt. We wanted to model influence observed among agents in a natural way so we developed a new model of interpersonal tie strength that was capable of measuring asymmetric relationships in a network based on the influence an agent had over another. We applied the model in synthetic and real-world networks and found that influence selection methods which did not ignore the influence they had over others, were the most efficient in propagating influence in synthetic networks and had the lowest misinformation propagation in the real-world networks.

Table of Contents

1 Introduction

Misinformation is a concerning issue that is increasingly prevalent with the recent emergence of online social media. The dissemination of misinformation in social-networks irrespective of its intent, is notably dangerous as it has the potential to provoke a large, often naïve social response. The problem is aggravated when both genuine and fake information circulate where the task of determining the authenticity of information is non-trivial particularly when previously reliable sources are initiators of misinformation.

The propagation of misinformation in online social-networks is accelerated by exploiting vulnerabilities in our social structures, notably the trust we place in our close contacts where we are likely to trust information presented to us regardless of its reliability if it's endorsed by our friends [23, 20]. Similarly, our extended exposure to trusted sources and types of content influence our motivation to propagate information and our ability to discern the reliability of the information we are exposed to [12, 7].

Current research [23, 36, 32] looks primarily at the characteristics and behaviour of misinformation however much is still unknown about how the dissemination of misinformation can be reduced. The gap in research ignores a very central problem that we investigate in this paper where we use an agent-based approach to model online social-networks and misinformation dissemination behaviour and exhaustively apply the model in synthetic and real-world networks to look at how the propagation of misinformation exploits the structural properties of a network and with optimizations of the model, how the propagation can be reduced.

Our agent-based model captures information dissemination using an influence adoption mechanism where the action of propagation is to influence and an influenced agent adopts the states of their influencer. Our agents maintain several states as attributes notably, their influenced and unauthentic states which describe the agent has been influenced and the reliability of the agent's information it's propagating. Both agent states are controlled during seeding where an initial randomly selected portion of the agent population adopt reliable and unreliable information. After the initial set of adopters are generated, agents run autonomously, trying to influence their neighbours and propagate their information until the entire network is influenced.

We introduce multiple agent decision making mechanisms which are used to simulate different online user behaviour patterns, where agents can make locally optimal decisions on which neighbour to influence as well as which neighbour to accept influence from. An agent's optimal choice is the evaluation of a ranking function where the method of ranking is determined by the system's rule set. The agent's choice of influence impacts how diffusion is carried out in the network where we look at centrality-based neighbour selection methods while an agent's choice of influencer affects how agents follow trends.

We propose a new model of interpersonal tie strength called the *Mutual Neighbour* model which measures the strength of asymmetric relationships observed in networks based on the influence a node has over another by evaluating the mutual frequency of neighbours between the nodes incident of a connection. The model is motivated by expectations of locality resilience where nodes that have greater influence over those adjacent, should be met with less resistance in the transmission of communication. The model responds well to common network properties including network density, clustering and heavy-tailed degree distributions. The Mutual neighbour model is applied in our agent-based model as the probability of influence in the model's probabilistic influence propagation mechanism.

The agent-based model is applied in a variety of synthetic networks generated from common random network generating models including the Erdős–Rényi random, Kleinberg small-world and Barabasi-Albert scale-free models. Additionally, we also apply the model on a real-world Facebook network consisting of over 4000 nodes that exhibits high clustering and low network density. We measure the influenced and unauthentic population proportions including their rates of change as well as the average number of steps influence propagates from a seed and also, the coverage of influence a seed has over the network

In our findings we observe that not only do neighbour selection methods which rank neighbours according to their centrality and factor in the probability of influence, propagate the least misinformation in the real-world networks but they are also the most efficient in propagating information in the synthetic networks in comparison to greedy centrality-based neighbour selection methods which do not factor in the probability of influence. These findings were also consistent when applying the different methods in synthetic networks with high clustering. Additionally, we find that the network size is a weak factor in misinformation propagation where the agent mechanism methods and network properties such as density, clustering and degree distributions were much larger contributing factors. Additionally, we find that misinformation rarely propagates beyond a few steps of the seed and large cascades that adopt misinformation contribute minimally to the average unauthentic population proportion.

The structure of this paper is as follows: In section 2 we introduce our Mutual Neighbour model of interpersonal tie strength. In section 3 we propose an agent-based model that captures influence behaviour among online-users in social-networks. In section 4 we outline our measures and experiment procedures as well as the experiment data used consisting of both synthetic and real-world networks where in section 5 we present our results which are discussed in section 6. Finally, in section 7 we conclude the findings of this paper and discuss our future work.

2 The Model

2.1 Mutual Neighbour Model

In this paper a new model of interpersonal tie strength named the *Mutual Neighbour* (MN) model is proposed, that measures the strength of connections based on the mutual frequency in locality of nodes incident of a connection. The motivation to design such a model is the expectation that the propagation of information in a network should be met with less resistance when the destination trusts the source, such an assumption is not unreasonable and is reconciled in recent literature that recognizes the value of strong ties in networks [25, 18]. The model provides a measure of tie strength that is captured entirely from the network topology while previous measures commonly need external factors to evaluate the strength of ties [20, 2].

The traditional notion of tie strength uses a loose definition of trust to describe relationships, the MN model extends this definition to include social influence and in the context of our framework this social influence expects that nodes are more likely to be *influenced* by those adjacent to them if they trust their influencer OR the influencer is *influential* in the network. Highly central nodes in the network are said to be influential where they have great influence over their contacts and as such, they are expected to be met with little resistance when trying to influence an adjacent node. The action of influencing in the network is for a node to adopt some information or ideas of their influencer, more specifically, in our framework this is to adopt the states of their influencer. Influence and more generally trust, are not always reciprocal [27] where one node may have more influence over another in a relationship and consequently, the MN model needed to be designed to measure asymmetric relationships.

The MN model intuitively computes the influence any given node u in a network has over an adjacent node $v \in N(u)$ as the ratio of mutual neighbours of the perspective node u and v (specifically the closed neighbourhood of u and open neighbourhood of v) to the number of connections incident of the adjacent node v where we have the following definition.

Definition 2.1. The *influence* any given node $u \in V$ in the graph $G = (V, E)$ has over another adjacent node v where $\{u, v\} \in E$ is the $influence_{u \to v}$

$$influence_{u \to v} = \frac{|N[u] \cap N(v)|}{degree(v)}$$

The closed neighbourhood of the perspective node u, $N[u]$ is used instead of the open neighbourhood such that it includes the node u itself, to resolve special cases in the network most notably, leaf nodes and more generally, the case in which two adjacent nodes u, v have no neighbours in common (except for themselves) such that, if the open neighbourhood was instead used then $influence_{u \to v}$ would evaluate to 0.

2.2 Applying the Mutual Neighbour Model

The Mutual Neighbour (MN) Model is applied when transforming a connected undirected graph $G' = (V', E')$ into our *influence network* which extends the structure of a weighted digraph $G = (V, E, f)$ where the MN model is the codomain in the edge weighting function f such that we have the following definition.

Definition 2.2. The *Influence Network* is a weighted digraph $G_{influence} = (V, E, f)$. Let $G' = (V', E')$ be the underlying graph of the influence network. It follows that:

$$\exists\ (u, v)\ \in E \wedge\ (v, u) \in E \Rightarrow \{u, v\}\ \in E'$$

The weighting function $f: E \rightarrow \mathbb{R}$ is the mapping of an edge $(u, v)\ \in E$ to the influence the node u has on v, that is

$$f(u, v) = influence_{u \rightarrow v}$$

Intuitively, the transformation process generates an influence network $G = (V, E, f)$ from an undirected graph where the set of nodes is the same and for each edge in the undirected graph incident on some nodes u, v, we generate two directed edges (u, v) and (v, u) where the weights of each edge are computed using the MN model that indicate the influence the source node of the edge has on the target node. The weights in the network are utilized in the frameworks probabilistic propagation mechanism to determine the probability of influence.

Neighbour Ranking Function

Agents in the network are limited in the number of neighbouring agents they can propagate influence to each time unit and due to the probabilistic propagation mechanism, propagation actions are not certain and as such, the choice of the local agent to influence is not so simple and it would be advantageous for agents to make an optimal choice based on local information. To allow agents to choose an optimal neighbour, we have a *neighbour ranking* mechanism which ranks neighbours based on a score and chooses the highest ranked neighbour to influence where the control is the *neighbour selection method* which assigns a score to each of the agent's non-influenced neighbours.

By changing the neighbour selection method we can control overall how agents in the network will propagate their influence. In this paper we propose a MN weighted degree neighbour selection method which is a compromise of degree while still considering the influence a node has over another such that the most central agents are not necessarily the optimal choice. This neighbour selection method scores the adjacent nodes v ∈ $N(u)$ of an agent u based on the product of their degree and the influence the agent v has on a given neighbour u (which is also the value of the weight $w(u, v)$ in the influence network) This function is shown in formula 2.2

$$\forall\ v\ \in N(u) \quad score_{prop}(v)\ =\ degree(v) \times\ influence_{u \rightarrow v} \qquad (2.2)$$

2.3 Centrality-based Neighbour Selection Methods

As well as MN weighted degree, there is also interest in examining other neighbour selection methods, in particular we look at common centrality measures including degree-centrality and eigenvector-centrality. The process of which, consists of computing the centrality[1] for each node in the network where nodes are ranked based on their centrality that is then used by an agents neighbour selection mechanism to determine which neighbouring agent to propagate influence to, such that in this case the neighbour with the highest centrality is chosen. The two introduced centrality-based models don't consider the MN weights as part of their ranking function however they remain restricted by the model's propagation mechanism such that, the action of influence propagation is still probabilistic whose probability of influence is determined by the edge weights in the influence network. The interest to include centrality-based neighbour selection methods in this paper was to see how the population proportions and diffusion properties changed in comparison to the MN weighted degree method if agents naïvely tried to propagate to their most central neighbour while only considering the network topology and ignoring how much influence they have over them.

Degree-Centrality

Degree-centrality is a simple and early centrality measure which computes how influential a node is in a network by the number of connections incident to it i.e. the centrality of any node u in a graph $G = (V, E)$ is its degree [15] as shown in formula 2.3.

$$C_{degree} = \deg(u) = |\{v \in V|\ \{u, v\} \in E\}| \tag{2.3}$$

To avoid double counting edges, we consider only the degree of nodes in the underlying network if the network is directed. The degree-centrality measure is similar to the MN weighted degree ranking function which also has a degree-centrality component however it takes into consideration the influence an agent has over its neighbours where the product of the neighbours degree-centrality and influence probability are used as the neighbours score such that the most central agents will not always be the most optimal choice to propagate influence. While using only the degree measure to score neighbours, the most central agent will always be chosen.

[1] Degree-centrality or Eigenvector-centrality. Only one centrality measure is calculated and active as the networks neighbour selection method at any given time.

Eigenvector-Centrality

Eigenvector-centrality differs from degree-centrality in that, it scores nodes that are connected to influential nodes higher than nodes with non-influential connections such that a node with high degree may not necessarily have high eigenvector-centrality, it could be that the a node with low degree has high eigenvector-centrality due to its influential connections in the network [9].

When computing the eigenvector-centrality of nodes in the network we first organize the underlying graph $G = (V, E)$ into an adjacency matrix A with $|V|$ rows and columns where entries A_{ij} = 1 if nodes i, j are adjacent otherwise $A_{ij} = 0$. Then solve equation 2 for the centrality vector $\vec{c}$

$$\vec{c} = \frac{1}{\lambda} A\vec{c} \text{ Such that } A\vec{c} = \lambda\vec{c} \tag{2.4}$$

While there may be many different λ values that solve for a non-zero centrality vector $\vec{c}$ however the largest eigenvalue and its corresponding eigenvector $\vec{c}$ are required by the Perron-Frobenius theorem [31]. To compute the largest eigenvalue and eigenvector the power-iteration algorithm shown in algorithm 2.3 can be used.

Algorithm 2.3 Power-Iteration

Input: Graph G = (V, E)
Output: Eigenvector normalized by magnitude for the largest eigenvalue

1: $n = |V|$
2: A = Adjacency matrix of G
3: I_n = nxn Identity matrix
4: $B = A + I_n$
5: $\vec{c} = (1, 1, \ldots 1)^T$, $\vec{c}.length = n$

7: **while** $\vec{c}$ is changing* **do**
8: $\vec{u} = B\vec{c}$
9: $k = \vec{u}_0$
10: $\vec{c} = \frac{1}{k}\vec{u}$
11: **end loop**

12: $\lambda = k - 1$
13: $\vec{c} = \frac{1}{||\vec{c}||}\vec{c}$

14: return $\vec{c}$

The returned vector $\vec{c}$ is the eigenvector for the corresponding largest eigenvalue. The eigenvector $\vec{c}$ is normalized by its magnitude $||\vec{c}||$. The eigenvector-centrality for any node i in the network is the ith entry in $\vec{c}$

3 Agent-Based Social Network Model

3.1 Modelling Influence in Social-Networks

An agent-based model (ABM) is a powerful modelling method that simulates complex networks using a system of autonomous entities called *agents* [8]. An agent can capture a real-world entity such as humans or organizations where they are adaptive, heterogeneous and capable of making decisions. Additionally, an agent is described by the attributes or states it maintains and an agent's behaviour with its environment, interactions and the decisions it makes are determined by a set of rules in the system. ABM's have had applications as early as 1970 with Shelling's model of segregation [33] and now more recently with modelling of human cognition [35] and large social networks [22, 34].

In this paper an agent-based model is used to simulate information sharing in online social-networks where agents in the network are modelled as online users. Presence of connections in the network describes a channel of communication between the agents incident of a connection that allows agents to share information to those adjacent to them while absence of a connection denies this communication among agents. An agents action of sharing information is synonymous with *influencing* where an agent that accepts information shared to them is influenced such that they adopt and more generally, endorse the information given to them. To reflect this, the behaviour of influencing in the model is extended such that agents now inherit the attributes of their influencer. To model influence in the network, the influence network $G_{influence} = (V, E, f)$ defined in section 2.2 is used and agents maintain their influenced state as well as influencer ID in their set of attributes. The influenced state is a binary flag to indicate if the agent has been influenced or not, while the influencer ID is a reference to the influencing agent. Additionally, a reference to the ID of an agents *influence tree* $T_{influence}$ is maintained where the ID references the associated tree's root.

Definition 3.1. An *influence tree* is the rooted tree $T_{influence} = (V, E, r_T)$ where for agents in the tree, $v \in V, influenced(v) = true$ and the root of the tree r_T is a seed in the influence network $G_{influence}$ such that r_T has the properties: $influencer(r_T) = \emptyset$ Additionally, an agent's parent in $T_{influence}$ for any agent $u \in V$, is its influencer $influencer(u)$

Intuitively, an influence tree is a rooted tree in the influence network that stems from an initially influenced agent (seed) where all agents that exist in an influence tree have adopted the states of the influence tree root agent. Additionally, it can be found that the influencer of an agent is its parent in the agent's influence tree. Moreover, if the reference to an agent's influencer ID does not exist but the agent is in an influenced state, then the agent itself is a seed and consequently the root of its own influence tree.

Maintaining influence trees helps analyse how much influence an agent has over the network and also how deep the influence propagates from the seed.

While the model does not have knowledge of the information *content* being propagated in the network, it can however identify the authenticity of the information such that we can observe misinformation being propagated in the network. To model this in the ABM, agents maintain as an attribute, the state of their information authenticity such that, an agent with an unauthentic state is an indication that the agent has adopted misinformation. The agent's information authenticity state is inherited by the agents they influence, such that an agent carrying misinformation will continue to propagate misinformation through the network and those agents that are influenced by the unauthentic agent do so also. Agents are not aware of their own or their influencer's authenticity state and it is not a factor when agents make decisions.

An agent's goal in the ABM is to influence all of its neighbouring agents and once this goal is achieved, the agent becomes inactive. The action of influence propagation to adjacent agents in the influence network is probabilistic where the probability of propagating influence from any agent u to another non-influenced adjacent agent v, is evaluated by $influence_{u \to v}$ and in the influence network this is the edge weight $w(u, v)$. Due to the probabilistic nature of this action, agents should make an optimal decision on which adjacent agent to influence where in section 2.2, the model's neighbour selection mechanism and multiple selection methods used to determine the optimal neighbouring agent to influence is discussed.

3.2 Agent Policies

Agents in the proposed ABM are able to make decisions on which neighbouring agents to propagate influence to, however it would be advantageous for neighbours to also be capable of deciding which neighbours to be influenced by. To model the additional decision making of choosing an influencer, a policy mechanism is introduced where agents maintain a list of *influence offers* at each time step and instead of an agent's influenced action using a first-come-first-serve method (an agent is influenced by the first agent to successfully propagate their influence), successful influencers are instead added to the targeted agent's influence offers. Each non-influenced agent with influence offers pending at the end of a time unit evaluates their influence offers to choose an optimal influencer by ranking the potential influencers based on the current network policy.

We were interested to see how particular trending behaviours observed in online social-networks affected the propagation of misinformation in a network where the motivation for policy methods was to look at how agents affected the network if they followed trends.

Trending-Source Policy Method

The trending-source policy expects agents to rank their influencers based on their accumulated propagation count in the network which is the frequency at which an agent has influenced others such that of the influence offers, the agent that has influenced the most agents, is the optimal choice.

Trending-Tree Policy Method

In this policy method, an agent's influence offers are ranked based on the size of their influence tree. The motivation of this policy method is that if an offering agent's influence tree has significant coverage in the network, it must be that belonging to this influence tree is advantageous such that of the influence offers, the agent that belongs to the largest influence tree is the optimal choice.

4 Methods

4.1 Network Models

In the field of network science the study of network modelling is the development of models that accurately capture characteristics of real networks. It is well-researched, particularly in the domain of sociology where modelling of social structures has been extensively studied with literature as early as 1932 [30]. More recently with the emergence of social media, mathematician's interests have been in modelling social networks [14] including contributions to statistical methods of analysis and modelling of complex network topologies such as small-world [38] and scale-free structures [5].

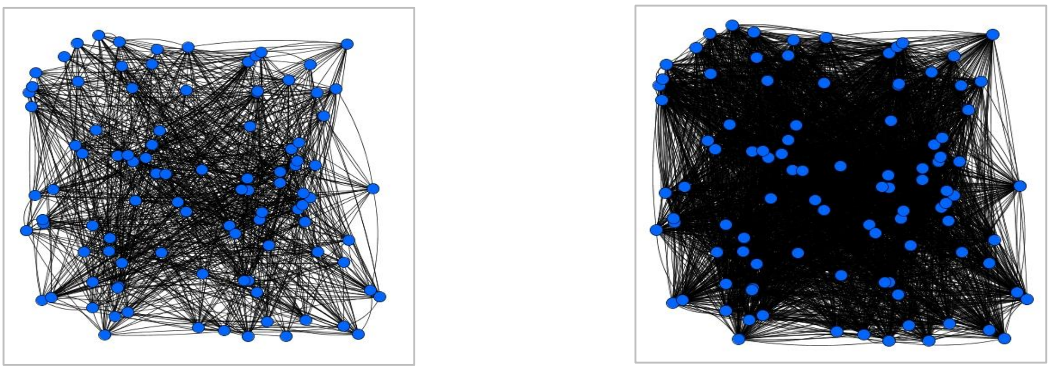
We look to use network models in this paper to simulate networks that our diffusion model can run on. Having a variety of network models with differing real-world network characteristics allows us to observe how the model behaves and performs in comparison to other models tested under particular simulated network environments and architectures. Experimenting on classical structures alone regardless of order[2], would be insufficient as they are not probabilistic models and follow a fixed architecture that would consequently limit our model and yield predictable results.

The experiments conducted in this paper make use of both generated synthetic networks and real-world collected networks. The synthetic networks were created from three random network models including Erdős–Rényi, Kleinberg and Barabasi models. Each model was capable of reproducing characteristics found in social networks where the Kleinberg model was able to capture 'small-world' and clustering behaviour while the Berbasi-Albert preferential attachment model was capable of generating scale-free networks whose architecture is widely observed particularly in social-networks [4]. The Erdos-Renyi model does not observe many real-world

[2] The number of nodes in a graph G = (V, E) such that order(G) = |V|

network characteristics as it lacks clustering, is not scale-free and has an unreasonable degree distribution, however it offers a truly random method of generating networks and serves as a useful baseline for our experiments. The networks generated are undirected as our mutual neighbour model takes an underlying graph as input and transforms the edges & the associated weights. Additionally, the model is limited to processing connected[3] graphs and as such the random network models must meet this connectivity requirement or be adjusted accordingly which became the case for the Erdos-Renyi model. Typically, the experiments generate networks with varying order and fixed model parameters, in the case that the network model parameters are changed then the order is fixed. The design of each random network model and their respective parameters are discussed in the subsequent sections.

The Erdős–Rényi (ER) model



(a) *(b)*

Fig. 4.1: An Erdős–Rényi network with 100 nodes and (a) low edge probability of $p = 0.1$, (b) high edge probability of $p = 0.5$

The ER model was first introduced in 1959 and is one of the earliest random graph models. Two variants of the model were proposed, the first by Paul Erdos and Alfred Renyi who introduced the $G(n, m)$ model [13] where a graph is randomly generated with n nodes and m edges from a possible $C(n, m)$ graphs and each edge is added with $\frac{1}{n}$ probability. The second variant by Gilbert [16] proposed the $G(n, p)$ model where a graph is generated with n nodes and there exists an edge from each node to another in the graph with fixed probability p. Unless otherwise stated, Gilbert's $G(n, p)$ variant of the ER model is used in this paper.

[3] There exists a path between each node in a graph

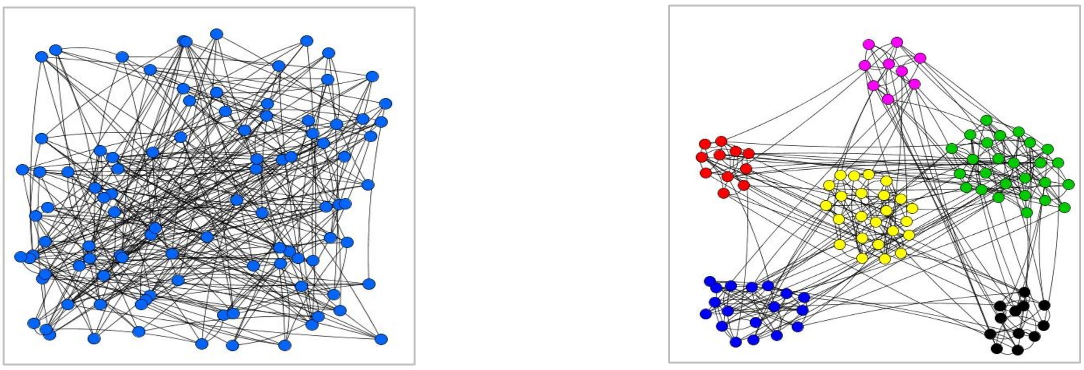
Algorithm 4.1 Extended ER

Input: Order n, Edge Probability p
Output: Connected undirected Graph with n nodes

Create $G = (V, E)$
Add n nodes to the graph G
for each $u \in V$ **do**
    Randomly select one node $v \in V$ where $v \neq u$
    $E \cup \{u, v\}$
    for each $u' \in V \cap u' \neq u \cap \{u, u'\} \notin E$ **do**
        $E \cup \{u, u'\}$ with probability p
    end for
end for

We find that increasing the edge probability p, increases a graph's density where at $p = 1$ we have a complete graph and an empty graph for $p = 0$. The ER model does not ensure a connected graph even for large values of p with the exception of $p = 1$, however generating complete graphs to guarantee connectedness may not be desirable furthermore, generating disconnected graphs is even less desirable as it violates our models requirements. To ensure connectedness we extended the ER model shown in Algorithm 4.1 by adding an additional step in the procedure before each node adds an edge to another node with probability p, where each node first connects with one other randomly selected node in the graph such that we have at least a connected graph before adding additional edges. In the experiments the edge probability p is fixed to 0.1, given that the extended model already produces a connected graph, values much larger than this yielded very dense networks.

Kleinberg model



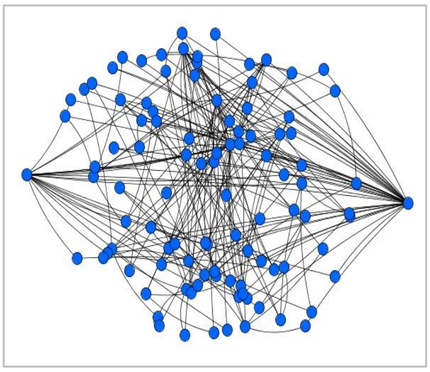
(a) *(b)*

Fig. 4.2: A Kleinberg network with 100 nodes and clustering exponent 2
(a) Network clustering not highlighted, (b) Network clustering displayed

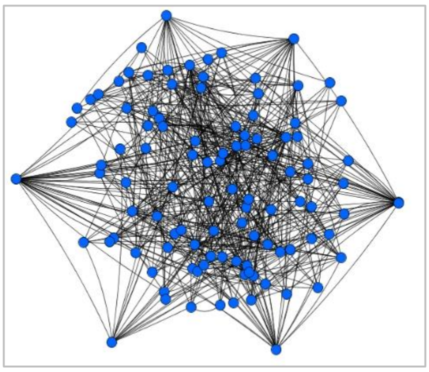
Small-world phenomena is the concept that we are all linked to each other not necessarily directly but through a string of acquaintances and was popularized by Stanley Milgram [29] where it was suggested that our societies structure is a small-world network. In terms of graph theory, a small-world network is a network where there may not exist an edge between each pair of nodes however each node can access another through a short path or hops more specifically, these networks have a low average path length and high clustering coefficient.

The Kleinberg model was introduced in 2001 [24] and proposed a greedy small-world routing algorithm. The model organizes a network of n nodes into an nxn lattice where each node is a point on the lattice. The nodes have local information of who their neighbours are in the lattice and as such a node v can make up to $|N(v)|$ local connections. Additionally, nodes have some acquaintances deeper in the lattice where the *lattice distance* between any two nodes is the number of steps in the lattice separating them. Nodes make long-range connections to other nodes in the lattice with probability $\frac{1}{ld(u,v)^r}$ where $ld(u, v)$ is the lattice distance from nodes u, v and r is the clustering exponent in the network. For small clustering exponent values nodes will be more likely to choose long distance nodes in the lattice than those in the local vicinity while large clustering exponents cause long range nodes to cluster closer in the lattice to the node. In [24] Kleinberg finds the optimal clustering exponent to be 2 which is the inverse square distribution where long-range connections are distributed evenly over all distances in the lattice. In this paper we fix the clustering exponent to the optimal value of 2. Additionally, we allow nodes to create connections to each of its neighbours in the lattice but are limited to one long-range connection to another node that is not a neighbour in the lattice with $\frac{1}{ld(u,v)^r}$ probability.

Barabási–Albert model



(a)



(b)

Fig. 4.3: A Barabasi-Albert network with 100 nodes and
(a) initial nodes $n_i = 2$, (b) initial nodes $n_i = 5$

The Barbasi-Albert model was proposed in 1999 [6] and is used for producing scale-free networks. Scale-free networks have a power-law or heavy-tailed degree distribution where there will exist *hubs* that are nodes in a network which are highly connected such that they have substantially larger degree than many other nodes in the network. Hubs in scale-free networks contribute the most to connectivity while in random-networks all nodes contribute approximately the same to connectivity. Many networks are regarded as scale-free such as social-networks, World Wide Web, protein and citation networks. The model generates scale-free networks by using a preferential-attachment mechanism to produce the power-law degree distribution where more central nodes are more likely to be connected to new nodes in the network than those not as connected. The model first generates a connected graph with n_i initial nodes where $\forall\ v\ \in V, deg(v) \leq\ 2$ then any additional nodes to be added in the network are connected to at most n_i nodes with probability $\frac{\deg(u)}{\sum_{v_i}\deg(v_i)}$ where $\deg(u)$ is the degree of an existing node and $\sum_{v_i}\deg(v_i)$ is the sum of degree's in the network. In experiments the number of initial nodes h n_i is fixed to 5% of the network order. Since the model uses the ratio of a nodes degree to sum of degrees as the connection probability, it's quite clear to see that new nodes will be more likely to connect to more central nodes than those with low degree and consequently the initial nodes in the connected network are more likely to become hubs than those nodes added later in the network as the initial nodes already have connections established.

4.2 Real-World Networks

The synthetic networks provide an efficient source to rigorously test and have the flexibility to control the network parameters as needed however they do not fully capture real-world networks and are not representative of actual data that is relevant and has meaning. We now look to include a real network from an external source to our set of networks that will strengthen our experiments as our data will now have context.

The real-network used in this paper comes from the Stanford large network dataset collection sourced from [28] where Facebook friend lists were collected from survey participants in the Facebook 'Social Circles' app during 2012. The dataset consists of 4039 nodes, 88234 edges. The network is connected with an average degree of 43, 1.1% graph density[4] and 0.605 average clustering coefficient[5].

The Facebook friendship circles data set was decided on as it most closely relates to the intended context and framework of this paper where we are interested in the behaviour of propagating influence and the spread of misinformation among users in a

[4] The ratio of edges to maximum possible edges in the graph that is computed with $\frac{2|E|}{|V|(|V|-1)}$

[5] Average clustering coefficient of all nodes in a graph where the clustering coefficient of an individual node is how complete the node's neighbourhood is.

social network so for this purpose, having a network of Facebook users is very beneficial.

4.3 Measures

In this section we cover what methods of measuring were used during experiments including the types of data and structures recorded. There were two main measuring components in this paper namely population proportions in the network and diffusion properties computed from the influence tree structures. Computing many of the measures was quite costly particularly for larger networks where repeated processing of large sets of nodes was occurring, so instead we opted for a dynamic programming approach where measures were recorded in hash tables and global counters then developed over each time unit during the agent influencing and polling processes. This approach imposed a larger memory footprint but yielded faster processing as measures could be obtained in constant time.

For measuring population proportions the interest was in recording the influenced and unauthentic population proportions including their rate of change from one time unit to the next. Additionally, the number of time units taken to influence a network entirely or the population proportions at a specific time unit was recorded.

When observing diffusion in the experiments we looked to measure how deep influence was propagating from the seeds and also how much influence a seed had over the network so for these measures the influence tree structures recorded were analysed for tree heights and sizes. Each of the measures recorded and their procedures are explained below.

A. Time Unit Count

The number of time units it has taken to influence a network such that at the max time unit there are no active agents and 100% of the network is influenced.

B. Influenced Population Proportion

The proportion of agents in the network with a positive influenced state P_{infl}. It is the ratio of influenced agents to total agents in the network $P_{infl} = \frac{|V_{infl}|}{|V|}$

C. Unauthentic Population Proportion

The proportion of agents in the network that are both influenced AND have an unauthentic state P_{unauth} are recorded at each time unit, that is the ratio of unauthentic agents to total agents $P_{unauth} = \frac{|V_{unauth}|}{|V|}$. We are typically interested in the unauthentic population proportion at the end of the diffusion process i.e. recorded unauthentic population proportion of the last time unit.

D. Population Proportion Rate of Change

The change in the population proportion from the previous time unit P_{t-1} to the current time unit P_t that is $P_t - P_{t-1}$ where we also compute the average rate of change in each population proportion which is $\frac{P_{tmax-}P_{tO}}{n}$ for n time units.

E. Influence Tree Height

The heights of each influence tree in the network are computed where trees stem from seed agents (such that the number of trees in the network is equivalent to the number of seeds where seeds are the roots of each influence tree). Each agent maintains their tree root agent and calculates their depth in the tree when influenced which is then compared with the respective tree root's recorded maximum depth and replaced if the recently influenced agents depth is larger than the depth recorded. An agent's tree depth is computed with algorithm 4.3 where each agent maintains reference to their influencer. Note that seeds and non-influenced agents have no influencer and as such they have depth 0 in the tree. The algorithm computes tree depth for an agent by recursively traversing the chain of agents back to their seed and counting each step.

Algorithm 4.3 treeDepth

Input: Agent x
Output: The tree depth of x; 0 if x is a seed

1: **if** x.influencer is null **then** return 0
2: **otherwise** return treeDepth(x.influencer) + 1

The height of a influence tree in the network is then the recorded maximum depth in the tree. We also compute the average tree height which is $\frac{\sum_{i=0}^{Sn} tree\_height(S_i)}{|S|}$ where the set S contains the seeds in the network.

F. Influence Tree Size

The influence tree size is the number of agents in an influence tree and gives an indication of how much influence any given tree has over the network. We compute the size of all trees in the network where each tree maintains its size and when an agent is influenced the agent becomes part of their influencer's influence tree (the influencer is a branch in a larger tree if it is not a seed) and the size of the associated influence tree is increased. We also record the average and maximum tree heights at each time unit.

4.4 Experimental Setup

The experiments conducted in this paper investigate the affect each of the model mechanisms had on the propagation of unauthentic information in both synthetic and real-world networks. The setup and parameters of each experiment are detailed in this section. Additionally, the general experiment process is illustrated and the details of the software written to run the experiments is also mentioned. Experiment results are presented in section 5.

All experiment tasks in this paper were run on the open source general graph modelling tool Graphi. The software was written specifically for this paper and also used in [26] as it allowed the integration of the complex agent-based model and as well as giving complete control over algorithm implementations that other software could not facilitate. The focus of Graphi was to create an interactive general-purpose graph modelling tool that was intuitive to use while still allowing complex tasks to be performed. The software features network simulating, state recording and playback, data views, plugins, interactive network display and is highly configurable while also supporting many common graph file formats.

Each experiment follows a similar setup process where initially a network model is chosen to generate a network[6]. Additionally, the neighbour propagation selection methods and policy methods need to be set. From the network generated 10% of the agents are randomly selected as seeds where each seed has the influenced state and of the seeds, a further 30% are randomly selected to adopt unauthentic states such that the initial unauthentic population proportion is 0.03. The generated network is then transformed using the mutual neighbour (MN) model where undirected edges are replaced with directed edges and associated edge weights for each edge are determined by the MN model. The initial network state at time unit = 0 is recorded including the computed measures[7] described in section 4.2. Next the diffusion process is ran where each active agent (initially the set of seeds) is polled and requested to propagate their influence to uninfluenced neighbours. Newly influenced agents are added to the set of active agents and the network state as well as the measures at the current time unit are recorded. The diffusion process repeats until the entire network is influenced or the process runs for the maximum number of time units which unless otherwise stated, is 100 time units. Experiments running on the synthetic networks that are generated from the random network models, use fixed model parameters defined in section 4.1 where unless otherwise stated, the Erdos-Renyi model has a

[6] In the case of the real-world network, the provided network is used

[7] Each time a network state is recorded, all measures mentioned in 4.2 are recorded

fixed edge probability of 0.1, the clustering exponent is fixed to 2 in the Kleinberg model and the number of initial nodes in the Barabasi-Albert model is 5% of the network order. The real-world network does not have any parameters that are controlled during experiments.

A. MN Weighted Degree Ranking Function Applied in Synthetic Networks

In this experiment the diffusion model with the MN Weighted Degree neighbour selection method is run on all three synthetic networks. Additionally, the policy mechanism uses the first-come-first-serve[8] method. The purpose of this experiment was to investigate how unauthentic propagation was affected if agents factored in both the network topology and the probability of influence in their choice of optimal influence.

B. Centrality-based Ranking Functions Applied in Synthetic Networks

The diffusion model instead uses the Centrality-based neighbour selection methods and runs them on each of the synthetic networks. For this experiment the degree-centrality and eigenvector-centrality ranking functions are experimented independently. The first-come-first-serve method is still used in the policy mechanism.

C. Trending Policies in Synthetic Networks

In this experiment we look at how trending behaviour affects unauthentic propagation in the three synthetic networks where the policy mechanism trending-source and trending-tree methods are tested instead of first-come-first-serve. The MN weighted degree is used as the neighbour selection method.

D. Increased Clustering in Synthetic Networks

In this experiment it was investigated how each of the neighbour selection method and policy mechanisms perform under varying degrees of clustering. The Kleinberg small-world network model was used to generate the networks where the lattice size was fixed to 23 and the clustering exponent parameter was changed in the range of [0.5, 5] incrementing 0.5 each time. First it was looked at how the MN model weights scaled on this network[9] with different clustering exponents then each of the different diffusion model mechanism methods was ran on the network with different clustering exponents.

---

[8] Agents with a first-come-first serve influence offer mechanism are influenced by the first agent to successfully influence them instead of adding successful influencers to a list and ranking them.
[9] The Kleinberg network with a lattice size of 23 was transformed using the MN model for different clustering exponents where the average edge weights generated by the MN model were recorded.

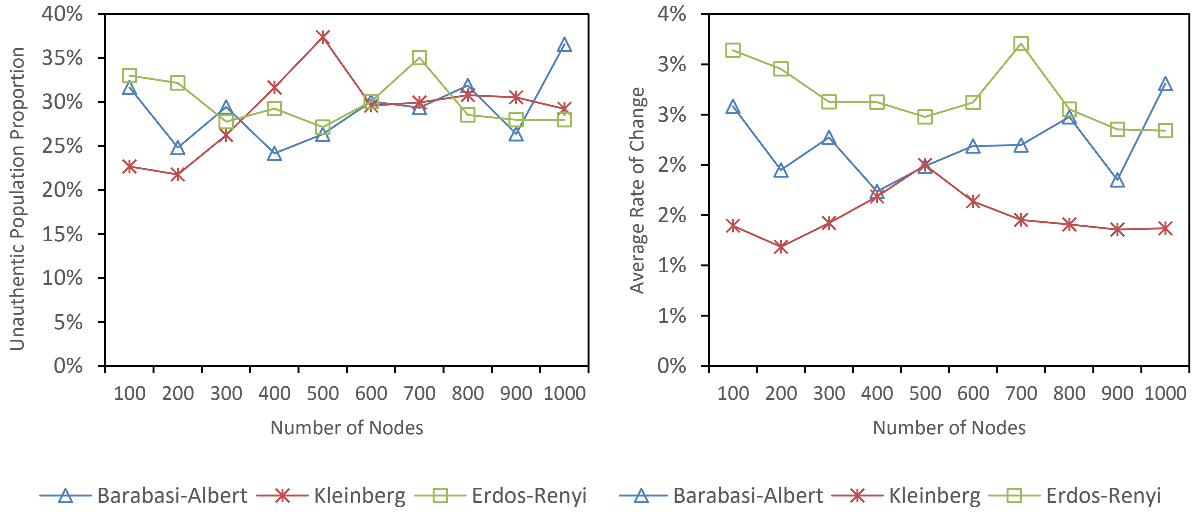
E. Applying the Model in Real-World Networks

We look at how all of the neighbour selection methods and policy methods perform on the provided real-world network. This network is detailed in section 4.2 and has fixed node and edge sets with no changeable parameters.

5 Results

The results of the experiments detailed in section 4.3 are presented as follows. It is noted that all experiment results are averaged over 10 trials where it was found the population proportions, rates of change and influence tree height measures observed the least variance while average and maximum influence tree size measures experienced the highest variance and weren't as useful during analysis as other measures. Experiments on synthetic networks have an order range [100, 1000] with fixed network model parameters (with the exception of the *increased clustering in networks* experiment which has fixed order and variable clustering exponent). Experiment measures and setup can be found in sections 4.3 and 4.4 respectively while experiment results are discussed in section 6.

A. MN Weighted Degree Ranking Function Applied in Synthetic Networks



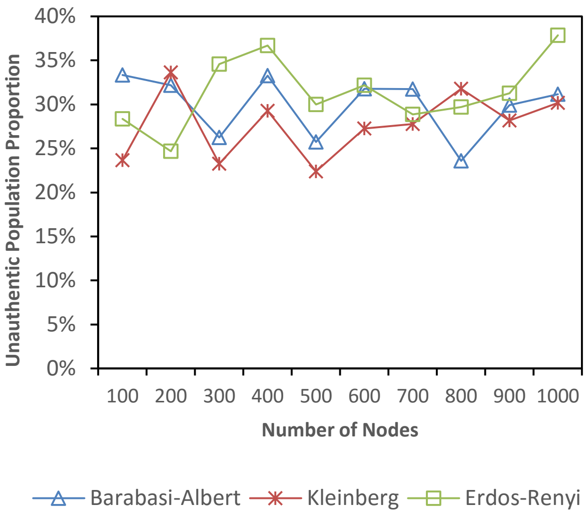
(a) *(b)*

Fig. 5.1: (a) Unauthentic population proportions of each synthetic network. (b) Unauthentic population proportion average rate of change in the synthetic networks.

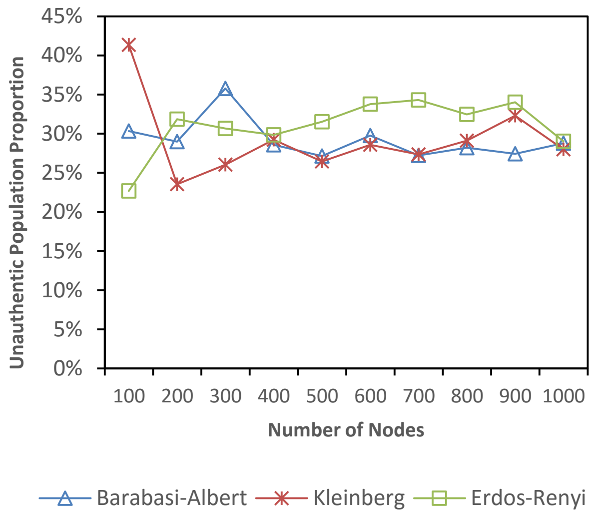
Experiment A uses the *MN Weighted Degree* neighbour selection method in the diffusion model that is run on each of the synthetic networks. The results for experiment A. are shown in figures 5.1a and 5.1b. It was found that the MN Weighted

Degree method running on the Kleinberg small-world model observed on average the least unauthentic population proportion at 29% with a 1.5% average unauthentic proportion rate of change while the Erdos-Renyi random model propagated on average the most at 30% with a 2.7% average rate of change. It is noted that the model performed very well on the Kleinberg networks for small sets of nodes but particularly poorly for the 500 node instance. While the Erdos-Renyi model did experience the highest unauthentic proportions out of the three models with the MN weighted degree function, its network influenced change rate of 9% was the highest of the three and on average it completed the influence diffusion process in only 10 steps where the Kleinberg was the least efficient and took on average 17 steps to influence the entire network. The deepest influence trees were found in the Kleinberg model however it also had the lowest maximum influence tree size of the three models.

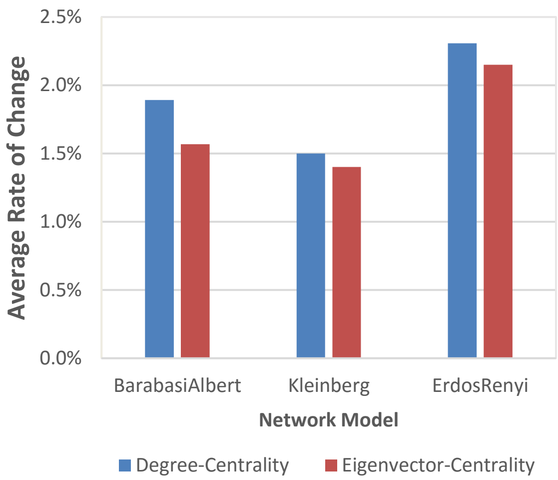
B. Centrality-based Ranking Functions Applied in Synthetic Networks



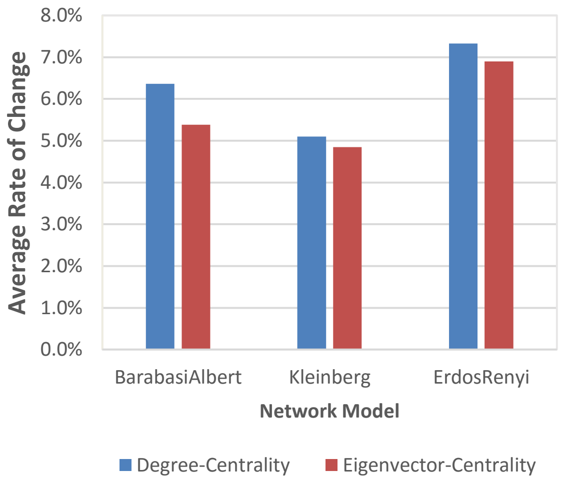
(a)



(b)



(c)

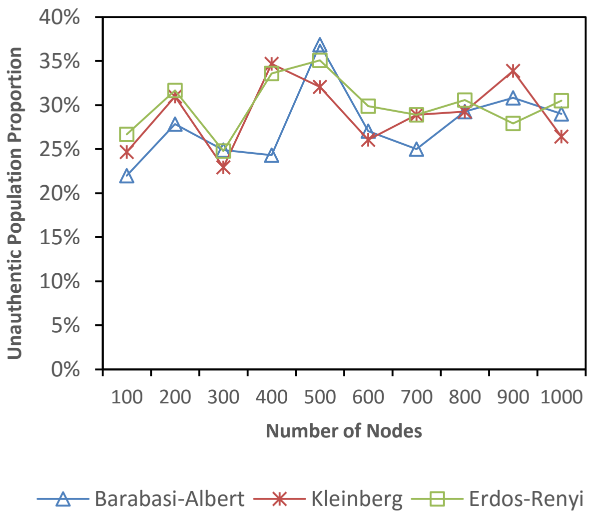


(d)

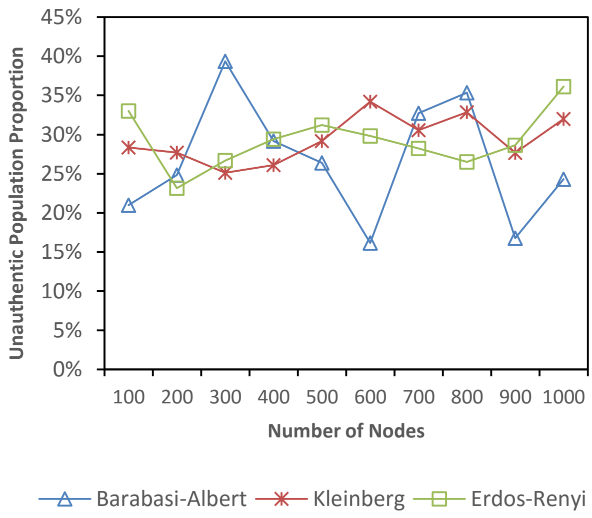
Fig. 5.2: Unauthentic population proportions of synthetic networks with (a) degree-centrality and (b) eigenvector-centrality neighbour selection methods. (c) The unauthentic population proportion average rate of change (d) influenced population proportion average rate of change.

Experiment B uses the degree-centrality and eigenvector-centrality based neighbour selection methods in the model on the three synthetic networks. Results for experiment B. are shown in figures 5.2a-5.2d. Of the two centrality-based ranking functions, the degree-centrality ranking function caused the highest average influence rate of change in all three network models however it also had the highest average unauthentic population proportion as well as influence rate of change in each of the networks. The Erdos-Renyi network model observed the highest unauthentic population proportion in both centrality-based ranking functions while the Kleinberg model had the lowest average unauthentic population proportion for degree-centrality and Barabasi-Albert recorded unauthentic population proportions were the lowest for eigenvector-centrality. The difference in average influence tree heights between the two centrality-based ranking functions was minimal where the Kleinberg networks had the highest average influence tree depths. Degree-centrality was more efficient in influencing the network for all three networks models however the difference in the average time units was at most 2 between the two centrality-based ranking functions.

C. Trending Policies in Synthetic Networks



(a)

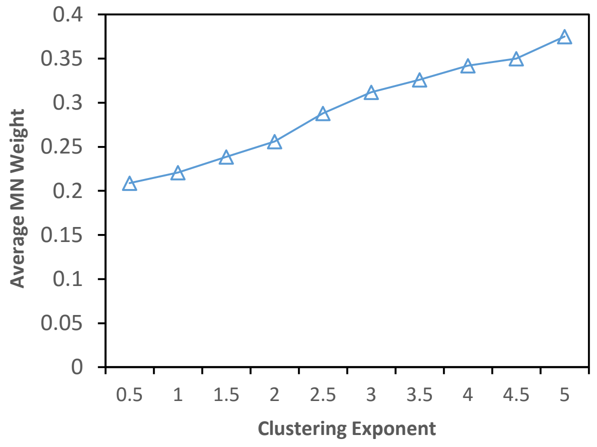


(b)

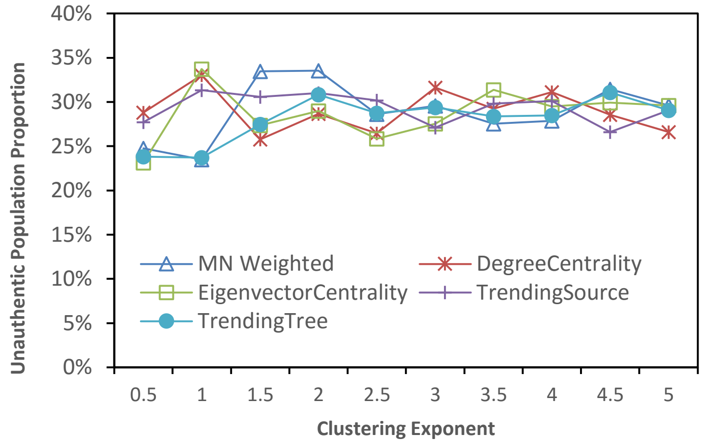
Fig. 5.3: (a) Unauthentic population proportions in synthetic networks with (a) trending-source and (b) trending-tree policy methods.

Experiment C changes the model's policy mechanism to use the trending-source and trending-tree policy methods. Each method is applied in the synthetic networks. Results are shown in figures 5.3a – 5.3b. In both the Barabasi-Albert and Erdos-Renyi networks, the trending-source policy propagates on average the most unauthentic information and has a marginally less average unauthentic population proportion in Kleinberg networks of the two policy methods. The difference in the networks average influence tree height and population rates of change was found to be very minimal between the two policy method results. Erdos-Renyi networks had the fastest influence propagation where the average time unit taken to influence the network was 12 with a 7.4% influence rate of change for both policies while Kleinberg was the least efficient where it took on average 18 steps for the trending-source and 19 steps for the trending-tree policies to influence the network. The Barabasi-Albert networks observed the most shallow influence trees while the Kleinberg networks had the deepest.

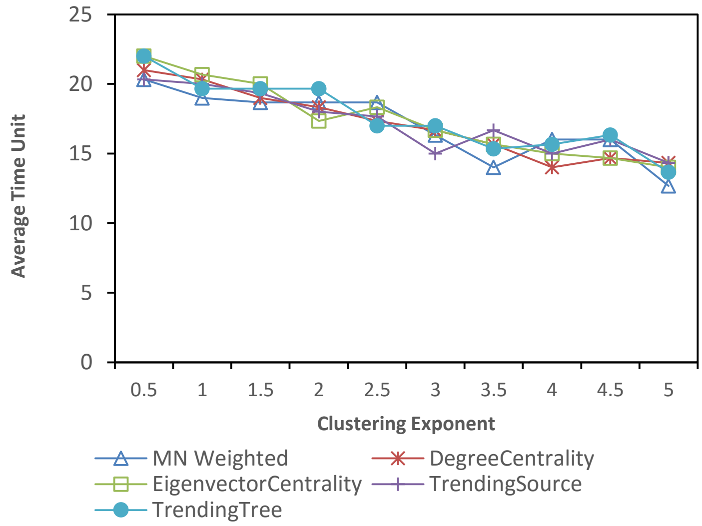
D. Increased Clustering in Synthetic Networks



(a)



(b)



(c)

Fig. 5.5: (a) The Mutual Neighbour weights in fixed size Kleinberg network with varying clustering showing a strong positive correlation. (b) Unauthentic population proportions in the Kleinberg network with varying clustering for each of the different diffusion methods. (c) Average time units of diffusion methods in the Kleinberg network with varying clustering exponents showing a strong negative correlation.

Experiment D runs all neighbour selection methods and policy methods on a Kleinberg network with 500 fixed nodes and different clustering exponent values. Results are shown in figures 5.5a – 5.5b. The average MN weights in the network responded well to increased clustering where it observed a strong positive correlation to the clustering exponent parameter with a correlation coefficient of 0.99. It was found that for the MN weighted degree method, the average unauthentic population proportion decreased by 21% for clustering exponents larger than the default clustering exponent (2) while the centrality-based neighbour selection methods observed an increased unauthentic population proportion where the degree-centrality method saw a significant 30% increase and eigenvector-centrality method 11% increase in unauthentic population proportion with larger clustering exponents.

All methods found an increased average influence rate of change and consequently reduced number of time units to influence the network for clustering exponents larger than 2. Average tree heights also observed a similar behaviour where methods experienced increased tree heights for larger clustering exponents.

E. Applying the Model in Real-World Networks

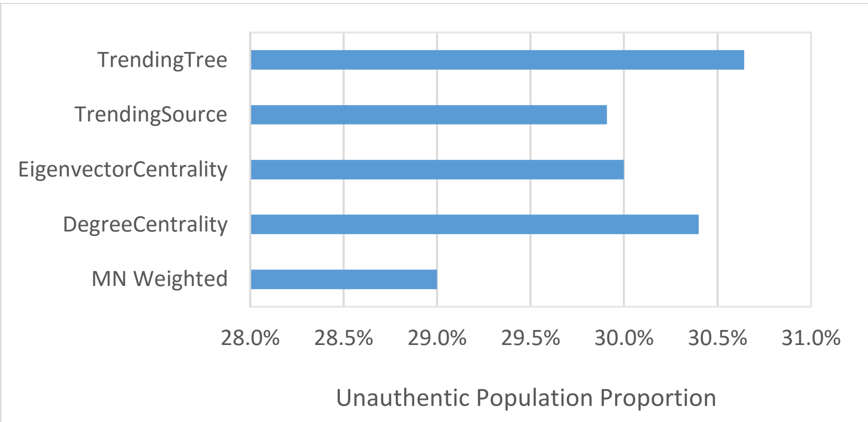


Fig. 5.4: Unauthentic population proportions for the Facebook network with different diffusion mechanisms active

Experiment E runs each of the neighbour selection methods and policy methods on the 4039 node, 88234 edge Facebook network. Results are displayed in figure 5.4. Out of all the diffusion mechanisms, the MN weighted degree propagation function performed the best on the real-world network where it observed the least unauthentic population proportion of 29% while the recorded unauthentic proportions of other mechanisms were all at or above 30% with the trending-tree policy method having the highest recorded unauthentic population proportion of 30.6%.

The real-world network had the largest number of nodes and was the densest network tested and as such, it had the highest recorded average time units where on average the diffusion process with each of the different methods took 30 steps to influence the entire network with the trending-source policy method being the least efficient in propagating influence in the network. The influence average rate of change with each method ranged from 2.5% - 4% which was much lower than the average influence rate of change recorded in the synthetic networks. The influence trees observed were shallow where the average tree height recorded across all methods was 2.6 with MN weighted degree experiencing the deepest trees.

6 Discussions

By applying the proposed mutual neighbour model of tie strength in our agent-based influence model, we have been able to observe how influencing behaviour affects the propagation of misinformation in both synthetic and real-world social networks where the contribution of this work was to provide a new understanding in how different sharing behavioural patterns exhibited in online users induces misinformation in social-networks. In the following section the findings of this work are discussed.

In experiments A-C each of the diffusion methods was applied in the Erdos-Renyi, Kleinberg and Barabasi-Albert synthetic networks with fixed model parameters and variable network sizes. It was found that across all tested methods, the Erdos-Renyi and Barabasi-Albert networks observed on average, the highest unauthentic population proportions. The high unauthentic population proportions found in the Erdos-Renyi networks was unexpected as it was hypothesised that the network model would experience lower unauthentic proportions as it had low clustering was not deliberately scale-free however, despite fixing a low wiring probability, the Erdos-Renyi networks experienced the highest density of the networks and while also having unpredictable degree distributions due to the probabilistic wiring component as well as its random selection where these factors were believed to be causing the increased unauthentic proportions. To make further sense of this, during initial parameter testing, both the average MN weights and unauthentic population proportion on a fixed 500 node Erdos-Renyi network observed a strong positive correlation with the connection wiring probability parameter such that high network density was found to be a contributing factor in the networks unauthentic population proportion. The high unauthentic proportions found in the Barabasi-Albert model was more expected due to the model’s heavy-tailed degree distribution such that if hubs in the network were initially selected with unauthentic states it was likely that they would contribute more to the unauthentic population proportion. Unauthentic population proportions measured in synthetic networks were found to be very weakly correlated with the size of the networks and while there was a noticeable increase in unauthentic proportions when comparing much smaller sized networks to larger networks however this observation was inconsistent and indicated the network size was not a strong factor in the propagation of unauthentic information in the synthetic networks but rather the network model parameters and the diffusion methods.

None of the diffusion methods applied in the synthetic networks stood out noticeably better than the others over all the networks where each method had strengths in a particular networks while observing weaknesses in others. In the Barabasi-Albert networks agents following trends in the network was found to be optimal where the policy methods notably tree-policy, observed the lowest unauthentic population proportions in comparison to the neighbour selection methods while the degree-centrality neighbour selection method experienced the highest unauthentic population proportion of the methods tested and this is expected due to its greedy degree selection method that is punished in a network with a heavy-tailed degree distribution. However in the Kleinberg networks, the degree-centrality neighbour selection method noticeably outperformed other methods as it was able to exploit the low-density and poisson degree distribution of the Kleinberg network. While in the Erdos-Renyi networks both centrality-based neighbour selection methods and in particular, degree-centrality again observed the highest overall unauthentic population proportions of the diffusions methods where it is clear that the greedy centrality-based neighbour selection methods do not behave well on networks with high-density where the methods using the MN weighted degree neighbour selection which factor in the MN weights, experienced noticeably less average unauthentic population proportion in comparison to the centrality-based neighbour selection methods.

The influence propagation efficiency of the diffusion methods in the synthetic networks was determined largely by the networks they were running on however, the MN weighted degree neighbour selection method was found to be the most efficient of the methods across all networks. It could be one’s expectation that centrality-based neighbour selection methods would be the most efficient of the methods tested as their ranking functions select the most central or influential neighbour and consequently have access to more connections faster than other neighbour selection methods and in recent studies [21], centrality-driven propagation was found to be very efficient however in the proposed model, the action of influence propagation is probabilistic where the probability of propagation is the influence an agent has over another as defined by the MN model, such that an agent which has less influence over its neighbour is less likely to propagate its influence so the low propagation efficiency of the centrality-based neighbour-selection is clear as the centrality-based methods select the most influential neighbours without factoring in the probability of influence while the MN weighted degree neighbour selection does.

A point of interest during our experiments, was to observe the influence tree structure measures recorded and examine how consistent the characteristics of the proposed diffusion model and the host of mechanisms introduced were to existing diffusion models where we looked at how much influence any given agent had over a network which was measured using the influence tree size measure as well as how many steps from the seed the influence jumps and was measured using the tree height. It was found that the average maximum influence tree depth (height) was very shallow where it was rare that influence propagated beyond 3 steps of a seed. In earlier literature [19, 37] a common expectation in the characteristics of diffusion was that a node's coverage stretched over large distances in the network exhibiting multi-step diffusion where large chains of referrals could be traced to their origin however such behaviour was not observed in our findings which were more consistent in recent studies [3, 1] which also finds that large multi-step diffusion is rare and propagation is contained within a few degrees of its origin. Additionally, the classical diffusion models [19, 37] suggest that these large cascades are rare and are often observed as *epidemic* events however when they do occur, they contribute the majority of the adopted population in the network. In our findings which are consistent with [17], these large cascades which were observed by the maximum tree sizes in the network were not only rare but they did not contribute the majority of the influenced population where it was found that at most their coverage was 10% of the influenced population.

The experiments investigating increased clustering in fixed size Kleinberg networks found that the average MN model weights of the influence network and the clustering exponent parameter observed a strong positive correlation and this result was reasonable as having a larger clustering exponent increased the probability of wiring closer on the lattice and this would reflect in the influence network where it would be more likely for nodes to have increased mutual neighbours with adjacent nodes and consequently a higher MN weight. It was also found that all diffusion methods on the fixed size Kleinberg network experienced increased influence propagation efficiency such that for each method, the average time unit and clustering exponent had a strong negative correlation moreover. This is consistent with recent studies [10, 11], where it was found that the propagation of social behaviour is farther and more efficient in clustered networks due to the strong social structures.

After finding positive results in the correlation of the average MN weights in Kleinberg networks and its clustering exponent parameter, we looked at applying each of the diffusion methods on the fixed size Kleinberg network and observing how the methods behaved with increased clustering. It was expected that with increased clustering, the methods would observe a noticeable increase in unauthentic population proportion, particularly for those methods that did not respond well to the initial experiments on the Kleinberg network. This hypothesis was true for the centrality-based neighbour selection methods where the degree-centrality experienced an average 30% increase in unauthentic population proportions for clustering exponents larger than the initially tested clustering exponent of 2 while the eigenvector-centrality similarly had an 11% increase.

Surprisingly, the methods using the MN weighted degree neighbour selection method all observed decreases in the average unauthentic population proportion for larger clustering exponents where the MN weighted degree neighbour selection method on its own had a significant 21% decrease in the average unauthentic population proportion and the source-policy policy method (also using the MN weighted degree neighbour selection method) also noticed a 10% decrease in the average unauthentic population proportion. This finding could be a noticeable contribution to current research as the MN weighted degree neighbour selection method has demonstrated that it propagates less misinformation in networks with increased clustering while others observe increased propagation of misinformation in highly clustered networks. It is acknowledged, that while the contributions of our findings in synthetic networks are limited as the generated networks are just an emulation of real-world networks, they do however simulate characteristics of real-world networks and as such there are applications for the presented findings in real-world networks.

The real-world network experiments applied each of the diffusion methods in a Facebook network which had low density and high clustering. The most notable of the real-world network experiment findings was that, of the methods applied, the MN weighted degree neighbour selection method recorded significantly lower unauthentic population proportion. This strengthens our findings of the increased clustering in synthetic networks experiment where the MN weighted degree neighbour selection method observed reduced propagation of unauthentic information with increased clustering while other neighbour selection methods experienced the opposite and is further evidence that the proposed MN weighted degree neighbour selection method is a promising optimization in reducing the propagation of misinformation in social networks.

7 Conclusions

In this paper we have developed an agent-based model that aims to capture the characteristics of online sharing users so that we can simulate misinformation propagation in synthetic and real-world social networks to further understand its behaviour and investigate methods of reducing it. In our experiments, we investigated how different diffusion mechanisms, network architectures and parameters affected the propagation of misinformation where we were particularly interested in the observed unauthentic population proportions. The primary contributions of this paper are as follows:

- We propose a new model of interpersonal tie strength that evaluates the strength of asymmetric relationships based on the influence a node has over another. The model has an intuitive design and responds well to heavy-tailed degree distributions, clustering and high network density. The model measures tie strength entirely from the network topology without the need for external factors sensitive to the network context in order to evaluate the strength of ties.

- We find that neighbour selection methods which factor in the probability of influence respond well to clustering in networks where the MN weighted degree neighbour selection method's unauthentic population proportion decreased noticeably with increased clustering in the synthetic Kleinberg network while the unauthentic population proportions of other methods increased significantly.

- We found that the proposed MN Weighted degree neighbour selection method's influence propagation was the most efficient of the methods tested in each of the synthetic networks.

- The network size was found to be a very weak factor in the propagation of misinformation where the network model parameters and diffusion methods were much larger contributing factors.

- We observed that misinformation rarely propagates beyond a few steps of the seeds where the average maximum influence tree depth was very shallow and averaged less than 3 steps of its origin in the network. Additionally, not only was large multi-step propagation found to be rare in experiments, the contribution of the largest influence trees to the influenced and unauthentic population proportions was minimal.

- The MN weighted degree neighbour selection method was found to observe significantly less unauthentic population proportion than all other methods in the real-world Facebook network with high network density and clustering.

Future work

In future work we look to further enrich the agent-based model to capture more complex behaviours of online users and social networks. Additionally, we look to explore new mechanism methods as well as optimizing existing mechanisms to reduce the propagation of misinformation. Possible areas of optimization that weren't considered during this paper could be in the initial seeding mechanisms where different seeding methods such as influence maximization could be applied

We also look to apply the model in more real-world networks where in this paper the model was applied in a Facebook friendship circle network but we identify the usefulness of other sources such as Facebook posting data and Twitter retweet data. In particular, networks which are known to have misinformation circulating would be interesting so that the unauthentic population proportions observed in simulations could be compared with those in the real-world network.

The proposed *Mutual neighbour* model of tie strength has a lot of potential for future applications where the model is very general and because it functions on most network architectures, evaluates using only the network topology and isn't constrained by external factors, it demonstrates a promising model of tie strength that could be applied in a variety of different domains for measuring the strength of asymmetric relationships in networks.

8 Acknowledgements

The author would like to give thanks to Dr. Jiamou Liu and Dr. Murray Black for their supervision of this paper. Thanks are also expressed to the Stanford network dataset collection for providing the Facebook network used in this paper.

9 References

[1] Adar, E., Adamic, L. (2005). Tracking information epidemics in blogspace. *IEEE/WIC/ACM International Conference on Web Intelligence*.

[2] Bakshy, E., Rosenn, I., Marlow, C., Adamic, L. (2012). The Role of Social Networks in Information Diffusion.

[3] Bakshy, E., Karrer, E., Adamic, L. (2009). Social Influence and the diffusion of user-created content. *In Proceedings of the tenth ACM conference on Electronic commerce. Association of Computing Machinery*, 325-334.

[4] Barabási, A., Bonabeu, E. (2003). Scale Free Networks. *Scientific American*, 50.

[5] Barabási, A. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74 47–97.

[6] Barabasi, A., Albert, R. (1999). Emergence of Scaling in Random Networks.

[7] Bessi, A., Coletto, M., Davidescu, GA., Scala, A., Caldarelli, G., Quattrociocchi, W. (2015). Science vs conspiracy: collective narratives in the age of misinformation.

[8] Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems.

[9] Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2 113–120.

[10] Centola, D. (2010). The Spread of Behavior in an Online Social Network Experiment. *Scienc*e, 329, 1194-1197.

[11] Centola, D., Macy, M. (2007) Complex Contagions and the weakness of Long Ties. *American Journal of Sociology,* 113, 702-734

[12] Ciampaglia, G., Flammini, A., Menczer, F. (2014). The production of information in the attention economy. *Scientific Reports,* 5.

[13] Erdos, P., Renyi, A. (1959). On Random Graphs I. *Publications Mathematicae*, 6, 290-297.

[14] Freeman, C. (2011). The Development of Social Network Analysis—with an Emphasis on Recent Events. *The SAGE Handbook of Social Network Analysis*, 26-40.

[15] Freeman, L. (1979). Centrality in Social Networks Conceptual Clarification. *Social Networks,* 1, 215-239.

[16] Gilbert, E. (1959). Random Graphs. *Annals of Mathematical Statistics*. 30, 1141-1144

[17] Goel, S., Watts, D., Goldstein, D. (2012). The Structure of Online Diffusion Networks. *Proceedings of the 13th ACM Conference on Electronic Commerce,* 623-639.

[18] Granovetter, M. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory,* 1, 201-233.

[19] Granovetter, M. (1978). Threshold models of collective behaviour. *American Journal of Sociology*, 83(6), 1420-1443.

[20] Granovetter, M. (1973). The Strength of Weak Ties. *The American Journal of Sociology*, 78, 1360-1380.

[21] Guimaraes, A., Viera, A., Silva, A., Ziviani, A. (2013). Fast Centrality-Driven Diffusion in Dynamic networks. *Proceedings of the 22nd International Conference on World Wide Web,* 821-828.

[22] Hamill, L., Gilbert, N. (2010). Simulating Large Social Networks in Agent-Based Models: A Social Circle Model.

[23] Jagatic, T., Johnson, N., Jakobsson, M., Menczer, F. (2007). Social Phishing. *Communications of the ACM,* 50(10), 94-100.

[24] Kleinberg, J. (2001). The Small-World Phenomenon: An Algorithmic Perspective.

[25] Krackhardt, D. (1992). The Strength of Strong Ties: The importance of Philos in Organizations.

[26] Li, L., Liu, J., Russell, K. (2017). What Becomes of the Broken Hearted? An Agent-Based Approach to Self-Evaluation, Interpersonal Loss, and Suicide Ideation. *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, 436-445.

[27] Manning, P. (1992). Erving Goffman and Modern Sociology.

[28] McAuley, J., Leskovec, J. (2012). Learning to Discover Social Circles in Ego Networks. *NIPS.*

[29] Milgram, S. (1967). The Small World Problem.

[30] Moreno, J. (1932). Application of the group method to the classification of Prisoners. *Group Psychoterapy: A Symposium*, 8, 15-39.

[31] Newman, M. (2006). The mathematics of networks.

[32] Nguyen, D., Nguyen, N., Thai. M. (2012). Sources of Misinformation in Online Social Networks: Who to suspect?

[33] Schelling,T. (1971). Dynamic Models of Segregation. *Journal of Mathematical Sociology*, 1, 143–186.

[34] Singer, H., Singer, I., Herrmann, H. (2009) Agent-Based Model for Friendship in Social Networks.

[35] Sun, R. (2006). Cognition and Multi-Agent Interaction: From Cognitive Modelling to Social Simulation.

[36] Vicario, D. M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G. Eugene, S., Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113.

[37] Watts, D. (2002). A simple model of information cascades on random networks. *Proceedings of the National Academy of Science*, 5766–5771.

[38] Watts, D., Steven, H. (1998). Collective dynamics of small-world networks.