Network structure influence on simulated network interventions for behaviour change

Badham, J., Kee, F., & Hunter, R. F. (2021). Network structure influence on simulated network interventions for behaviour change. *Social Networks*, *64*, 55-62.

Published in:
Social Networks

Document Version:
Peer reviewed version

Network structure influence on simulated network interventions for behaviour change

Jennifer Badham[a,*], Frank Kee[a], Ruth F Hunter[a]

[a] *Centre for Public Health, Queen's University Belfast, Institute of Clinical Sciences, Block A, Royal Victoria Hospital, Belfast BT12 6BA, UK*

Abstract

We simulated diffusion of behaviour change over fifteen real-world networks with seven network interventions under both simple and complex contagion. We found that structural network properties affect both the diffusion outcome and the relative effectiveness of the different interventions, with confounding effects that were inconsistent with results expected from mathematical analysis. These results suggest that comprehensive studies are needed to identify the effects of structural properties on diffusion in real-world networks. Further, researchers attempting to identify the effect of individual properties must measure a range of properties to avoid incorrect attribution.

Keywords: Network interventions, Network diffusion, Centrality, Degree distribution, Clustering coefficient, Geodesics

*Corresponding author.

Email addresses: `research@criticalconnections.com.au` (Jennifer Badham), `f.kee@qub.ac.uk` (Frank Kee), `ruth.hunter@qub.ac.uk` (Ruth F Hunter)

Background

Network diffusion models represent a broad range of dynamic processes where active nodes in a network are able to activate neighbouring nodes. Applications include the propagation of disease, information, and behaviour, with 'active' representing infection, awareness and adoption respectively.

Adoption of new practices has been of research interest in several diverse disciplines since the early 1900's, starting with anthropology and rural sociology (Katz et al., 1963; Rogers, 2003). These early studies were strongly empirical, analysing the way in which practices spread through space and time. With the involvement of communication and marketing researchers, the next group of studies focussed on the mechanisms by which diffusion occurs. The seminal study of hybrid corn adoption (Ryan and Gross, 1943, 1950) demonstrated the importance of interpersonal factors in the spread of a new practice. This study found that acceptance was a slower process than awareness with a different balance between communication mechanisms; "salesmen were credited with informing the majority of the operators, but neighbors were credited with convincing them" (Ryan and Gross, 1950, pg 685). While these diffusion studies were occurring in independent disciplines, there was substantial overlap in their methods and results, which eventually converged as 'diffusion of innovations' (Katz et al., 1963; Rogers, 2003). That is, the insights concerning different communication mechanisms supported a general theory about adoption of new ideas or behaviours.

At around the same time, the emerging discipline of social network analysis (or sociometry, in contemporary terminology) was developing methods to identify those people within a community who are influential at an interpersonal level (Moreno, 1934). These methods included data collection instruments and analysis tools (Freeman, 2004, and references therein). Sociometricians were also demonstrating that interpersonal influence did not necessarily reside with people in positions of influence or of high status (Stewart, 1947; Jennings, 1950). These sociometric studies used degree centrality to measure opinion leadership, even while recognising that approach as problematic (Jennings, 1950). In response, Katz proposed a centrality index that included discounted in-degrees of out-neighbours "to study influence, transmission of information, etc." (Katz, 1953, pg 39), implicitly linking diffusion and influence.

These ideas converged in the now classic study of the diffusion of a new drug in four cities (Coleman et al., 1957). This study combined empirical information recording when each doctor first prescribed the drug (diffusion of the innovation) with comprehensive collection of information about the advice, discussion and friendship networks amongst relevant doctors (sociometry). No attempt was made to trace conversations about the drug, but these data were used to demonstrate that neighbours in the advice and discussion networks started prescribing at around the same time. The pilot study for this research used different analysis techniques and stated a similar result as "drug adoptions on any particular date are more frequent among doctors who are in direct sociometric contact with others who have already adopted the drug, than among doctors who lack such contact" (Menzel and Katz, 1955, pg 348). This pilot study identified three communities based on network structure, and each community had fast adoption within the community but at different times than the other communities, suggesting that this structure supported three distinct diffusion processes of behaviour change.

Theorising about the mechanism for transmission of collective behaviour, Granovetter (1978) introduced the widely used mathematical model of thresholds, where a person adopts the behaviour once some proportion of a group has already adopted. A key insight from this model is the sensitivity of the diffusion outcome to variations in the distribution of personal thresholds. Valente (1996) adapted this model to the personal influence perspective of the diffusion of innovations literature, proposing the proportion adopted within a person's ego-network as the relevant mechanism. In addition to providing a new analysis technique for empirical diffusion studies, the threshold model also described a mechanism for simulation or, for certain problems, mathematical analysis (such as the analysis of conditions under which cascades occur in Watts, 2002).

Of particular relevance to our study, Valente's focus on the ego-network added to potential sources of variation. In particular, even where all people have the same threshold, diffusion is sensitive to variations in degree because such heterogeneity also varies the amount that a single adopted network neighbour contributes to the proportion adopted. This contrasts with the collective behaviour result where identical thresholds lead to all or none adoption. It also opened research questions concerning the effect of the network itself on diffusion. As real-world networks (almost always) have different properties from each other and empirical studies confound network effects with individual effects such as variation in thresholds, mathematical analysis and simulation studies are an appropriate way to investigate such network effects.

Despite these difficulties, some researchers have been able to examine the effect of network structure experimentally. Most notably, Centola (2010) controlled network visibility for an online health community in six trials that each compared a clustered network (lattice) with a random network (rewired lattice) where each participant had identical degree and was randomly assigned a network position. Diffusion was initiated with a randomly selected participant informing neighbours about a health forum, with each new registration to the forum triggering further messages. In this study, diffusion was higher in the clustered networks for all six trials. This result provided evidence that multiple neighbours are needed to adopt to drive diffusion, as well as demonstrating the effect of network clustering on the diffusion outcome.

Separately, mathematicians had been developing models with a different diffusion mechanism. These are based on the categorical models of epidemiology, where people are classified according to their disease state as susceptible (not yet contracted the disease), infectious (contracted the disease and able to transmit it to others) or removed (no longer active in the epidemic process due to recovery with immunity or death) (Kermack and McKendrick, 1927). The model variants are denoted by the first letter of the included states, with SIR and SIS (where recovery does not confer immunity) the most frequently studied (see comprehensive review at Pastor-Satorras et al., 2015). In the simplest version, an infected person has a fixed probability of transmitting the infection to any other person they come in contact with. The network equivalent is a fixed probability of transmission along any edge connecting an infected and susceptible pair (Keeling and Eames, 2005). In the context of behaviour change, SI models can be used to represent the transmission of information or awareness.

In simulation studies, these two diffusion mechanisms are commonly referred to as complex and simple contagion respectively (following Centola and Macy, 2007). In simple contagion, any active node has some probability of infecting any inactive neighbours. In complex contagion, an inactive node is activated if the proportion of network neighbours that are currently active exceeds the relevant threshold. Versions with the same probability (simple) or same threshold (complex) for all nodes are the simplest implementation, but there are many variants of these contagion processes including, for example, limiting the period during which an active node is able to influence its inactive neighbours.

The effect of network structure differs between the two contagion mechanisms (Pastor-Satorras et al., 2015). With simple contagion, all nodes in the same component as the starting node(s) will eventually be activated due to the ongoing probability of transmission, but saturation will be faster with chains of high degree nodes. In contrast, complex contagion transmission can stall before saturation and, with identical thresholds, cascades occur through chains of low degree nodes. For example, a node with degree of one is activated at any threshold once its neighbour is activated. These differing effects have been studied both mathematically and through simulation.

Isolating first the effect of degree, consider the case where all edges connect randomly selected pairs of nodes so there is no degree correlation and the clustering coefficient is determined by density. For a lattice with all nodes of degree k, cascades will occur with simple contagion if the probability of transmission exceeds $1/k$, and cascades will occur with complex contagion if the threshold is lower than $1/k$. Degree heterogeneity reduces this critical value, increasing the speed of simple contagion (at least initially) and decreasing the proportion of the network reached with critical contagion (for details see Pastor-Satorras et al., 2015, including exact results for specific degree distributions).

Two other properties have attracted some attention - assortativity (or degree correlation) and clustering coefficient. With simple contagion, mathematical analysis shows that assortativity and clustering facilitate diffusion, confirmed with simulations (Pastor-Satorras et al., 2015; Newman, 2003). The effects are less clear under complex contagion (Guilbeault et al., 2018). Assortativity increases the likelihood of cascades occurring and the proportion of the network reached (Dodds and Payne, 2009; Payne et al., 2009). For clustering coefficient, however, results are limited to unrealistic networks with a clustering coefficient dependent on degree (Ikeda et al., 2010) or, to avoid confounding with assortativity results, identical degree for all nodes (Hackett et al., 2011). Even within these limited studies, increased clustering both increases and decreases the reach of the diffusion process at different value ranges. These results also depend on the specifics of the diffusion mechanism with, for example, clustering inhibiting diffusion under limited duration simple contagion(Newman, 2003).

While these results concerning the effect of network structure on a diffusion process are informative, they assume that the diffusion is initialised with a single node or a very small set of active nodes. Instead, public health behaviour interventions are initialised with a substantial proportion of the target population. There are many ways to use networks to enhance behaviour interventions, with four general approaches formalised under the term 'network interventions' (Valente, 2012). The 'individuals' approach uses network data to identify an initial set of people who are then expected to promote the intervention. This application of social network analysis was envisaged from the outset, with Moreno (1942, pg 304) confidently claiming that if high degree individuals "are won over to the idea, the balance of the population will almost automatically become infused with the necessary understanding and enthusiasm for the idea". While using opinion leaders (most commonly nominated, or degree centrality) is a common selection method, other centrality measures are also used (Valente, 2012; Hunter et al., 2019). The 'segmentation' approach selects some community within the network as the starting group. The other two approaches stimulate activity within the network to reinforce the desired behaviour ('induction') or encourage changes in the network ('alteration') to increase desirable connections or decrease undesirable ones.

Several simulation studies have compared diffusion under different choices of initial participants (or seeds) (including Aral et al., 2013; Zhang et al., 2015; Beheshti et al., 2017; Badham et al., 2018; van Woudenberg et al., 2019; Badham et al., 2019; Valente and Vega Yon, 2020). While results have been mixed, these studies have found that selection methods that preferentially recruit high degree seed participants generally lead to greater or faster adoption in comparison to random selection over several assumed behaviour adoption mechanisms and networks.

Two of these network intervention simulation studies (van Woudenberg et al., 2019; Valente and Vega Yon, 2020) also examined the effect of network structure. van Woudenberg et al. (2019) focussed on density and centralisation using 26 networks constructed from six school class relationships. The behaviour change mechanism in this simulation was based on a physical activity intervention and adjusted a continuously measured activity score rather than activating a binary choice. Each time step, an agent's behaviour value moved toward the weighted average of its neighbours' behaviour values, provided the difference was not too large. Simulations were initialised by increasing the behaviour values of the starting nodes by 17%, with the average change in behaviour value after 365 time steps used as the outcome measure. The study did not find an effect associated with density or with betweenness or closeness centralisation. However, larger changes in simulated behaviour were associated with higher degree centralisation, regardless of the centrality measure (degree, closeness or betweenness) used to select the initial group. Valente and Vega Yon (2020) focussed on threshold distributions and seed selection, but also included some network metrics. In particular, they found that higher modularity (or a stronger community structure) inhibited diffusion for complex contagion.

In this study, we combine elements from all these strands to explore the interaction between network structure and intervention effectiveness with both simple and complex contagion. We add to the small evidence base concerning the association between network properties and contagion outcomes generally. Additionally, we examine the relationship between network properties and different interventions. Fifteen real-world school friendship networks are used that display diverse degree distributions and other properties of interest. Consistent with both common practice in health behaviour network interventions and previous simulation studies, we assess seven network interventions that focus on degree and are primarily drawn from the individuals approach. The simulations apply fixed probability (simple) or threshold (complex) to model idealised versions of the separate mechanisms of awareness and behaviour change processes respectively.

Methods

We analyse data generated in a simulation study described elsewhere (Badham et al., 2019), with the simulation and network data available online at `https://osf.io/kjv4f`. That study used networks constructed from within class friendship nominations collected as part of the 2016 wave of the Wellbeing in Schools Survey (Davison et al., in prep). Of the 87 survey networks, only 17 met the completeness criteria for simulation, with at least 20 students and 80% of the students making nominations of which at least 80% were identifiable. The networks were symmetrised and isolates were removed. Simple and complex contagions were simulated over those 17 networks, starting from 15% of the network as selected by seven interventions.

In our study, we used these simulation data to investigate the contribution of different network properties in explaining differences in relative effectiveness of the interventions, as well as to the diffusion outcome regardless of intervention. While the original study used 17 networks, two (numbered 4 and 12) are disconnected. As our study uses proportion activated as a measure of contagion effectiveness and compares this measure between networks, we did not use the simulation data for the disconnected networks.

The networks are described in Tables 1 and 2. Eleven structural properties were examined, focussing on degree, clustering and geodesics because of their expected relevance to diffusion. Degree and clustering coefficient have been studied mathematically and with simulations and are known to affect diffusion, as described above. Geodesics (or shortest paths) are relevant to diffusion as they measure the number of successful transmissions needed to move across the network. As it measures the number of shortest paths that pass through a node, betweenness provides information about the potential for alternate paths that diffusion could take. We examine both the scale and variability of these properties, informed by the known importance of degree variability. Note that we have used the Gini coefficient as the measure of variability because it is robust to differences in network size (unlike centralisation measures) and makes no assumptions about the shape of the distribution of the property values (Badham, 2013). Extensively used as a measure of income inequality, the Gini coefficient of the degrees in a network is the normalised expected difference between the degrees of two randomly selected nodes in the network (and similarly for other network properties).

Table 1: Network properties[1] (Degree related).

Network	Nodes	Edges	Comm	M Degree	G Degree	Assortativity
0	22	84	4	7.6	0.22	0.21
1	27	132	2	9.8	0.14	0.14
2	21	94	3	9.0	0.19	-0.04
3	22	115	2	10.5	0.13	0.03
5	21	112	2	10.7	0.12	-0.15
6	22	81	3	7.4	0.22	0.20
7	33	157	3	9.5	0.19	-0.01
8	28	112	3	8.0	0.17	-0.08
9	17	64	2	7.5	0.10	0.26
10	24	89	3	7.4	0.21	0.36
11	20	89	2	8.9	0.13	0.07
13	26	106	4	8.2	0.28	0.47
14	20	61	4	6.1	0.19	0.03
15	24	111	2	9.3	0.15	0.46
16	29	101	3	7.0	0.26	0.41
Min	17	61	2	6.1	0.10	-0.15
Mean	23.7	101	2.8	8.4	0.18	0.16
Max	33	157	4	10.7	0.28	0.47

[1] Comm indicates number of communities identified with the Louvain algorithm (Blondel et al., 2008). M Degree and G Degree denote mean degree and Gini ceofficient of degree respectively.

An agent-based model was used to simulate a simple or complex contagion process initialised with 15% of the network (rounded up) as active nodes. Agent-based modelling is "a computational method. . . to experiment with models composed of agents that interact within an environment" (Gilbert, 2008, pg 2). The agents in this model are extremely simple, with no characteristics other than activation status. They occupy the nodes in the network and the environment for each agent is the activation status of the agents at adjacent nodes. All decisions are completely reactive as the agents are activated according to the rules of simple or complex contagion.

There are seven interventions included in the model, of which three randomly select starting participants: *Random Uniform* does so with equal probability and provides a baseline for comparison, *Friend of Random* has each uniformly randomly selected node select one of its neighbours and therefore introduces a bias toward higher degree nodes (Feld, 1991), and *Random by Degree* explicitly biases toward higher degree nodes to represent an outsider's attempt

Table 2: Network properties[1] (Path related).

Network	M CC	Transitivity	M Geo	Diameter	G Geo	G Btw
0	0.61	0.57	1.7	4	0.24	0.55
1	0.71	0.69	1.8	4	0.25	0.55
2	0.63	0.58	1.5	3	0.22	0.52
3	0.72	0.69	1.5	3	0.23	0.56
5	0.70	0.65	1.4	2	0.21	0.49
6	0.74	0.72	1.9	5	0.27	0.66
7	0.65	0.59	2.0	5	0.25	0.56
8	0.69	0.62	1.9	5	0.24	0.64
9	0.91	0.88	1.7	3	0.30	0.83
10	0.52	0.58	1.9	4	0.25	0.43
11	0.69	0.66	1.6	3	0.25	0.55
13	0.64	0.67	2.3	6	0.33	0.71
14	0.77	0.66	2.2	5	0.31	0.71
15	0.68	0.69	1.7	3	0.24	0.50
16	0.68	0.62	2.7	6	0.32	0.74
Min	0.52	0.57	1.4	2	0.21	0.43
Mean	0.69	0.66	1.8	4.1	0.26	0.60
Max	0.91	0.88	2.7	6	0.33	0.83

[1] M CC is the mean (local) clustering coefficient, with transitivity the global clustering ceofficient. Geo indicates geodesic, the shortest path between a pair of nodes, with M Geo, Diameter and G Geo denoting mean, maximum and Gini coefficient respectively. G Btw is the Gini coefficient of node betweenness.

to select the highest degree nodes. Two interventions, *High Degree* and *Community Leaders*, select high degree nodes across the network and distribution between communities respectively. These represent different ways of identifying opinion leaders, or the 'individuals' approach in network interventions (Valente, 2012). The *Community* intervention uniform randomly selects within a single community, representing the 'segmentation' approach. Finally, the *Persuasive* intervention selects randomly from throughout the network, but then attributes a higher influence to those initial seeds, intended to represent the 'induction' approach. These interventions are described further and justified in the study where the simulation data were generated (Badham et al., 2019).

Each simulation was initialised with the nodes selected by the intervention set to active, and all other nodes to inactive. For simple contagion simulations, each active node has some probability of activating each of its inactive neighbours. That probability is a parameter of the experiment, is identical for all nodes and constant over time. If the simulation is for the *Persuasive* intervention, the starting nodes are assigned a slightly higher probability (adding 0.2) of activating their neighbours. For complex contagion simulations, each inactive node calculates the proportion of its neighbours that are active and is itself activated if that proportion is at least the required threshold (an experimental parameter). For the *Persuasive* intervention, each starting node contributes twice in the proportion calculation. The simulations stopped naturally, either when all nodes had been activated (simple) or when no additional nodes would be activated in the next time step (complex).

From the previous study (Badham et al., 2019), the values of 0.7 for probability (simple contagion) and 0.5 for threshold (complex contagion) generate greater differentiation between simulation results, and these are the experimental settings for the results reported in this article. Other simulation results were available in the dataset and were checked for consistency: probability of transmission of 0.4 and 1, and threshold of 0.4 and 0.6.

Overall, there were 630 combinations of model settings (15 networks by 7 interventions by 2 diffusion types by 3 parameter settings for each contagion). The two interventions that select highest degree nodes are deterministic and always select the same initial nodes, except where there are draws in ranking. All others include randomness in the selection. Once the starting nodes are identified, complex contagion is deterministic and simple contagion is stochastic. A larger number of simulations were run for stochastic settings. The experimental design is summarised at Table 3.

Table 3: Experimental design: simulation parameters and number of runs.

Parameter	Values
Network	15 undirected, adapted from WiSe friendship nominations
Intervention	2 deterministic: High Degree, Community Leaders
	5 stochastic: Random Uniform, Random by Degree, Friend of Random, Community, Persuasive
Seed group size	15% of network, rounded up
Contagion	Simple, with transmission probability 0.4, 0.7 or 1
	Complex, with threshold proportion of 0.4, 0.5 or 0.6
Repetitions	5 for deterministic interventions and complex contagion
	100 for stochastic interventions and complex contagion
	100 for deterministic interventions and simple contagion
	1000 for stochastic interventions and simple contagion
Total simulations	256,950

The simulation stops when no further diffusion can occur. For simple contagion, this occurs when every node in the network is active. Therefore, one measure of relative effectiveness is the number of time steps to saturation (duration). For complex contagion, the simulation ends when all the inactive nodes have too few active neighbours. Therefore, these simulations can be compared using the proportion of active nodes achieved (penetration). In addition, we used the proportion of active nodes at one or two time steps (1-hop reach, 2-hop reach) as effectiveness measures. Each measure was calculated as the mean over all simulations (5, 100 or 1000) with the same simulation settings.

The simulation model was developed in NetLogo (Wilensky, 1999), a specialist agent-based modelling language, and its BehaviorSpace tool managed the batch simulations. All analysis was performed using R (R Core Team, 2015), particularly the 'igraph' package for network extraction and property calculation (Csardi and Nepusz, 2006), 'dplyr' package for analysis (Wickham and Francois, 2016), and 'ggplot' package for visualisation (Wickham, 2009).

Results

There are two related questions concerning the way in which network structure interacts with network interventions. The first concerns absolute effectiveness: do certain network properties enhance or inhibit diffusion? The second concerns differential impact: if so, are certain network interventions more or less influenced by the network properties? The effect of each network property can be seen by ordering the networks by the property of interest in a plot of effectiveness. The first question can be considered with plots that compare the effectiveness measure (for example, with colour), and the second with plots of the rank compared to other interventions of the measure. Property associations can be then observed as patterns (such as consistent colour changes) from left to right.

With the large number of simulation setting combinations and network properties, to avoid repetition, we present selected results in detail and refer to supplementary materials to support more general claims. The initial analysis focusses on duration for simple contagion with 0.7 transmission probability and penetration for complex contagion with 0.5 threshold.

From Figure 1 (left panel), high density is associated with faster saturation with simple contagion over all interventions, though the relationship may not be true for specific pairs of networks. This pattern is expected as higher density is equivalent to a greater number of edges for the probabilistic transmission to exploit. The most interesting pattern is for *High Degree*; this intervention is relatively more effective (ranks higher) for networks with higher density (Figure 1, right) and low ranked elsewhere, although there are inconsistencies such as low density network number 8 where *High Degree* is the most effective intervention. Other interventions show no density related pattern except to shift in response to the change in *High Degree*.

The opposite pattern occurs under complex contagion, with penetration for higher density networks lower than for low density networks for all interventions (Figure 2). The difference is strongest for the *High Degree* intervention, which is relatively effective at low density, but ineffective at higher density, with lower penetration than even the baseline of uniformly random initial seeds.

Another property of interest is degree heterogeneity or centralisation, measured using Gini coefficient. From the discussion in the introduction, heterogeneity is expected to enhance diffusion under simple contagion and inhibit it under complex contagion (Pastor-Satorras et al., 2015). We found the opposite effect in our simulations, with higher values for the Gini coefficient of degree associated with slower diffusion under simple contagion (Figure 3) and a larger

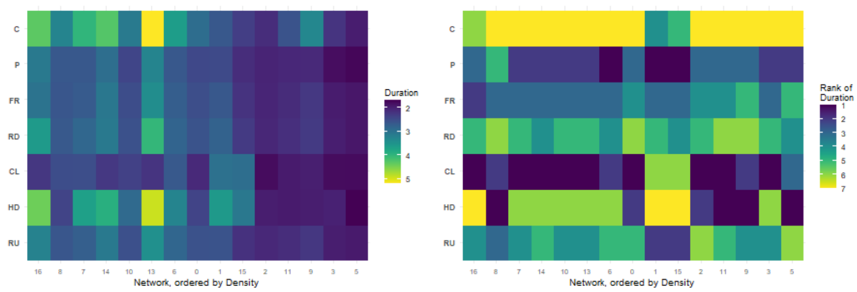

Figure 1: Effect of network density on absolute (left) and relative (right) effectiveness of interventions (measured with duration) under simple contagion with probability of transmission of 0.7. In both panels, bluer (darker) indicates faster saturation. Density of networks increases from left to right. Interventions are (top to bottom): Community, Persuasive, Friend of Random, Random by Degree, Community Leaders, High Degree, Random Uniform.

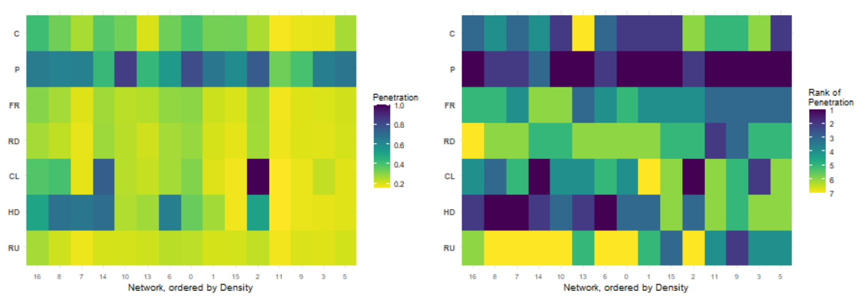

Figure 2: Effect of network density on absolute (left) and relative (right) effectiveness of interventions (measured with penetration) under complex contagion with threshold of 0.5. In both panels, bluer (darker) indicates greater effectiveness. Density of networks increases from left to right. Interventions are (top to bottom): Community, Persuasive, Friend of Random, Random by Degree, Community Leaders, High Degree, Random Uniform.

proportion of activated nodes under complex contagion (Figure 4). The likely explanation for this difference is confounding between structural properties; the correlation between density and Gini coefficient of degree is -0.78, so the effect of degree variation is overwhelmed by the larger effect of density. The patterns were consistent across interventions, however *High Degree* was relatively effective compared to other interventions for higher degree heterogeneity in the complex contagion simulations.

We did not observe any other patterns that suggested a relationship between a network property and the outcome of a contagion process over all interventions. However, the *High Degree* and, to a lesser extent, *Community Leaders* interventions are relatively effective for networks with longer paths under complex contagion, but not for simple contagion. This relationship occurs for both mean geodesic (see Figure 5) and maximum geodesic.

All of the described patterns also occurred where effectiveness was measured with 2-hop reach and over all transmission probabilities (simple) or thresholds

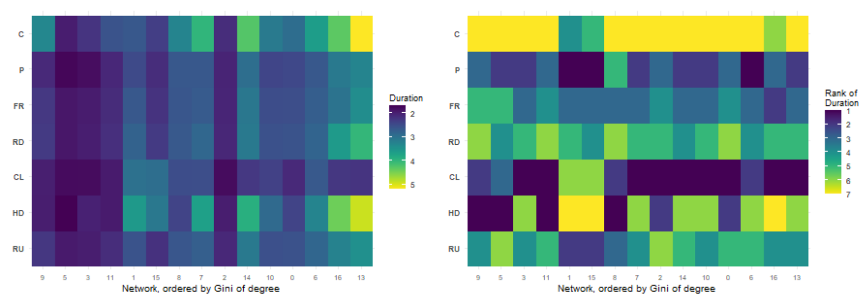

Figure 3: Effect of degree heterogeneity on absolute (left) and relative (right) effectiveness of interventions (measured with duration) under simple contagion with probability of transmission of 0.7. In both panels, bluer (darker) indicates faster saturation. Gini index of degree increases from left to right. Interventions are (top to bottom): Community, Persuasive, Friend of Random, Random by Degree, Community Leaders, High Degree, Random Uniform.

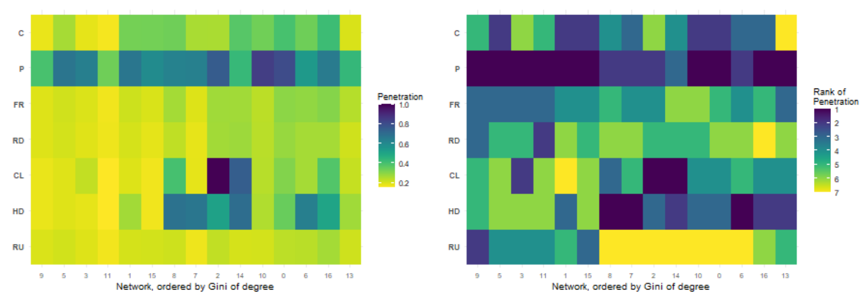

Figure 4: Effect of degree heterogeneity on absolute (left) and relative (right) effectiveness of interventions (measured with penetration) under complex contagion with threshold of 0.5. In both panels, bluer (darker) indicates greater effectiveness. Gini index of degree increases from left to right. Interventions are (top to bottom): Community, Persuasive, Friend of Random, Random by Degree, Community Leaders, High Degree, Random Uniform.

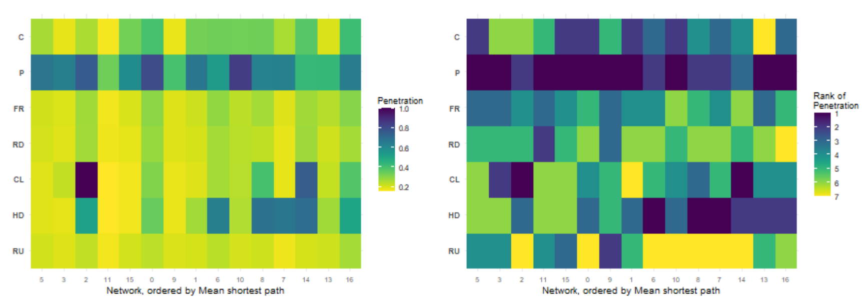

Figure 5: Effect of mean geodesic on absolute (left) and relative (right) effectiveness of interventions (measured with penetration) under complex contagion with threshold of 0.5. In both panels, bluer (darker) indicates greater effectiveness. Mean geodesic increases from left to right. Interventions are (top to bottom): Community, Persuasive, Friend of Random, Random by Degree, Community Leaders, High Degree, Random Uniform.

(complex). However, results were inconsistent for 1-hop reach. Full sets of charts are available from [Supplementary materials or online].

There is potential for substantial confounding between network properties. In addition to the high correlation between density and Gini coefficient of degree for the fifteen networks already noted, there are several other pairs of properties where any relationship between the property and simulation results could be obscured (see Table 4). For example, Gini coefficient of degree is strongly positively correlated with mean geodesic, as well as the negative correlation with density. Consider simple contagion, to the extent that longer paths overall are associated with longer paths from the initial nodes, a larger mean geodesic will increase the duration and may be the underlying cause of the observed pattern for degree heterogeneity.

Table 4: Correlations between network structural properties

Property	1	2	3	4	5	6	7	8	9	10
1 Nodes										
2 Density	0.59									
3 M Degree	0.12	0.73								
4 G Degree	0.39	-0.78	-0.62							
5 Assortativity	0.04	-0.46	-0.52	0.55						
6 M Clustering	-0.38	0.08	-0.27	-0.34	0.05					
7 Transitivity	-0.45	0.27	0.09	-0.42	0.09	0.86				
8 M Geodesic	0.51	-0.85	-0.66	0.73	0.51	0.04	-0.11			
9 Diameter	0.60	-0.91	-0.60	0.82	0.36	-0.13	-0.22	0.90		
10 G Geodesic	0.08	-0.56	-0.72	0.52	0.59	0.31	0.33	0.84	0.72	
11 G Betweenness	-0.10	-0.37	-0.61	0.25	0.22	0.75	0.62	0.57	0.49	0.79

Discussion

The first question considered asked whether certain network properties enhance or inhibit diffusion. The results show that the answer is yes, with structural properties related to degree showing the strongest patterns. These simulations are inconclusive, however, in the sense of being able to identify the effect of specific network structural properties on diffusion as even the most obvious patterns included some networks with inconsistent results. Further, there was confounding between the effects of different structural properties.

While this study simulated contagion over only fifteen networks, the networks were all collected in similar populations: school based, same age group and the same friendship relation. While this similarity could be expected to lead to networks with similar structure, the networks in the study differed over multiple properties with some properties showing a large range in values (see Tables 1 and 2). Consequently, there were no subgroups of networks that were sufficiently similar to isolate the effect of specific properties. Furthermore, the confounding between properties led to results that were inconsistent with expected outcomes based on mathematical analysis of simplified networks.

This suggests that more systematic analyses are required to uncover the effect of network structure on diffusion. The most theoretically sound approach would generate sets of networks that are socially realistic and differ from each other in only one property. These networks could then be used to support a detailed simulation study. However, this requires algorithms that generate networks that can control multiple property values simultaneously. Such algorithms are not available and are theoretically difficult due to the interdependence of structural properties. A more feasible approach is to undertake a similar simulation study as presented here, but on a much larger scale. As many studies are now collecting network data in schools, particularly for health interventions, such an approach is possible if sufficient of these networks are made public and collated. It is also possible that the effect of structural differences would be smaller in larger networks if subnetworks that enhanced or inhibited diffusion offset each other. Future work could repeat this study using networks collected, for example, in behaviour interventions conducted at the school year (rather than individual class) level.

Of course, even completely controlled simulation studies exclude some factors that would enhance or inhibit real-world diffusion. Most obviously, people vary in their personal thresholds or susceptibility to influence (Valente, 1996), though this could be accounted for in simulations if required. To the extent that personal characteristics that are associated with influence coincide with characteristics that are associated with network position (or forming of relevant relationships), these characteristics would also confound the effect of network structure on real-world diffusion.

The second question considered whether certain network interventions are more or less influenced by the network properties. The results show that relative effectiveness of network interventions also depends on the structure of the network and that this dependence is strong enough that network interventions that are relatively effective over some networks are less effective than even random selection of initial active nodes in other networks. This has important implications for behaviour interventions. Network interventions in schools to promote healthy behaviour typically select high degree students as their initial participants (Hunter et al., 2019) and such behaviour is expected to be diffused through a complex contagion process (Valente, 1996). From Figure 4, such an approach may be counterproductive where the network has little variation in degree.

As this study suggests that general guidance as to the ‘best’ network intervention is not possible, (at least) two paths are available to researchers who wish to optimise their network intervention. The first is to collect network information for their specific population and simulate different network interventions, assuming some appropriate diffusion mechanism, and selects an intervention approach accordingly. This would extend the precision prevention paradigm (Gillman and Hammond, 2016) from individual characteristics to social characteristics.

The second is to consider other criteria in selecting the network intervention. For example, the initial participants could be identified using the *Friend of Random* intervention as it balances high degree with practicality because it does not require network data collection (as was done in Kim et al., 2015). Alternatively, the network questions could ask about relationships that are directly relevant to behaviour change (such as advice seeking) rather than generic influence like friendship, in the expectation that high degree is beneficial. Finally, the study could use a network intervention that is not based on starting individuals, such as peer encouragement designs (Eckles et al., 2016) with additional specific communication over all existing network links (induction in the taxonomy of Valente, 2012).

These results also have implications for other studies intended to compare the effect of specific network properties using real-world networks on diffusion generally or to compare incentives. It is not sufficient to measure the property of interest and compare outcomes, ascribing differences to some relationship between the outcome and that property. A range of other potentially relevant properties must also be measured, to assess alternative explanations. At a minimum, diffusion studies should report degree, clustering and geodesic information for included networks, both scale (such as mean) and variation.

Conclusions

We found that structural network properties affect both the diffusion outcome and the relative effectiveness of the different interventions. The properties related to different aspects of degree showed the strongest effect, but were inconsistent with the effects expected from mathematical analysis of simplified networks. One likely explanation is that the much richer set of properties in real-world social networks confound each other, with some enhancing and others inhibiting diffusion. That is, the randomness assumptions in mathematical approaches and the unrealistic structure of synthetic networks compromise the capacity to apply the results.

The results from this study with richer network structures suggest that comprehensive studies are needed to identify the effects of structural properties on diffusion in real-world networks if general advice is to be developed about what intervention approach is most appropriate for different networks. As a separate issue, studies that attempt to isolate the effect of specific structural properties must measure a range of properties to ensure any observed influences are not attributed to the wrong property.

References

Aral, S., Muchnik, L., Sundararajan, A., 2013. Engineering social contagions: Optimal network seeding in the presence of homophily. Network Science 1, 125–153. doi:10.1017/nws.2013.6.

Badham, J., Kee, F., Hunter, R.F., 2018. Simulating network intervention strategies: Implications for adoption of behaviour. Network Science 6, 265–280. doi:10.1017/nws.2018.4.

Badham, J., Kee, F., Hunter, R.F., 2019. Effectiveness variation in simulated school-based network interventions. Applied Network Science 4. doi:10.1007/s41109-019-0168-6.

Badham, J.M., 2013. Measuring the shape of degree distributions. Network Science 1, 213–225. doi:10.1017/nws.2013.10.

Beheshti, R., Jalalpour, M., Glass, T.A., 2017. Comparing methods of targeting obesity interventions in populations: An agent-based simulation. SSM - Population Health 3, 211 – 218. doi:https://doi.org/10.1016/j.ssmph.2017.01.006.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008, P10008. URL: http://stacks.iop.org/1742-5468/2008/i=10/a=P10008.

Centola, D., 2010. The spread of behavior in an online social network experiment. Science 329, 1194–1197.

Centola, D., Macy, M., 2007. Complex contagion and the weakness of long ties. American Journal of Sociology 113, 702–734.

Coleman, J., Katz, E., Menzel, H., 1957. The diffusion of an innovation among physicians. Sociometry 20, 253. doi:10.2307/2785979.

Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. InterJournal Complex Systems, 1695. URL: http://igraph.org.

Davison, J., Lloyd, K., Dunne, L., Connolly, P., in prep. Longitudinal overview of adolescent wellbeing: findings from the wellbeing in schools survey .

Dodds, P.S., Payne, J.L., 2009. Analysis of a threshold model of social contagion on degree-correlated networks. Physical Review E 79. doi:10.1103/physreve.79.066115.

Eckles, D., Kizilcec, R.F., Bakshy, E., 2016. Estimating peer effects in networks with peer encouragement designs. Proceedings of the National Academy of Sciences 113, 7316–7322. doi:10.1073/pnas.1511201113.

Feld, S.L., 1991. Why your friends have more friends than you do. American Journal of Sociology 96, 1464–1477. doi:10.1086/229693.

Freeman, L.F., 2004. The Development of Social Network Analysis: A Study in the Sociology of Science. Empirical Press, Vancouver, Canada.

Gilbert, N., 2008. Agent-Based Models. Sage Publications, Los Angeles.

Gillman, M.W., Hammond, R.A., 2016. Precision treatment and precision prevention. JAMA Pediatrics 170, 9. doi:10.1001/jamapediatrics.2015.2786.

Granovetter, M., 1978. Threshold models of collective behavior. American Journal of Sociology 83, 1420–1443. doi:10.1086/226707.

Guilbeault, D., Becker, J., Centola, D., 2018. Complex contagions: A decade in review, in: Lehmann, S., Ahn, Y.Y. (Eds.), Complex Spreading Phenomena in Social Systems. Springer International Publishing, pp. 3–25. doi:10.1007/978-3-319-77332-2_1.

Hackett, A., Melnik, S., Gleeson, J.P., 2011. Cascades on a class of clustered random networks. Physical Review E 83. doi:10.1103/physreve.83.056107.

Hunter, R.F., de la Haye, K., Murray, J.M., Badham, J., Valente, T.W., Clarke, M., Kee, F., 2019. Social network interventions for health behaviours and outcomes: A systematic review and meta-analysis. PLOS Medicine 16, e1002890. doi:10.1371/journal.pmed.1002890.

Ikeda, Y., Hasegawa, T., Nemoto, K., 2010. Cascade dynamics on clustered network. Journal of Physics: Conference Series 221, 012005. doi:10.1088/1742-6596/221/1/012005.

Jennings, H.H., 1950. Leadership and isolation; a study of personality in interpersonal relations. 2nd ed., Longmans, Oxford, England.

Katz, E., Levin, M.L., Hamilton, H., 1963. Traditions of research on the diffusion of innovation. American Sociological Review 28, 237. doi:10.2307/2090611.

Katz, L., 1953. A new status index derived from sociometric analysis. Psychometrika 18, 39–43.

Keeling, M.J., Eames, K.T.D., 2005. Networks and epidemic models. Journal of the Royal Society Interface 2, 295–307.

Kermack, W.O., McKendrick, A.G., 1927. Contributions to the mathematical theory of epidemics - i. Proceedings of the Royal Society 115A, 711–721.

Kim, D.A., Hwong, A.R., Stafford, D., Hughes, D.A., O'Malley, A.J., Fowler, J.H., Christakis, N.A., 2015. Social network targeting to maximise population behaviour change: A cluster randomised controlled trial. The Lancet 386, 145–153. doi:10.1016/s0140-6736(15)60095-2.

Menzel, H., Katz, E., 1955. Social relations and innovation in the medical profession: The epidemiology of a new drug. Public Opinion Quarterly 19, 337. doi:10.1086/266584.

Moreno, J.L., 1934. Who shall survive?: A new approach to the problem of human interrelations. Nervous and Mental Disease Publishing Co. doi:10.1037/10648-000.

Moreno, J.L., 1942. Sociometry in action. Sociometry 5, 298. doi:10.2307/2784969.

Newman, M.E.J., 2003. Properties of highly clustered networks. Physical Review E 68. doi:10.1103/physreve.68.026121.

Pastor-Satorras, R., Castellano, C., Mieghem, P.V., Vespignani, A., 2015. Epidemic processes in complex networks. Reviews of Modern Physics 87, 925–979. doi:10.1103/revmodphys.87.925.

Payne, J.L., Dodds, P.S., Eppstein, M.J., 2009. Information cascades on degree-correlated random networks. Physical Review E 80. doi:10.1103/physreve.80.026125.

R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Rogers, E.M., 2003. Diffusion of Innovations. 5th ed., Simon & Schuster.

Ryan, B., Gross, N., 1950. Acceptance and diffusion of hybrid corn seed in two Iowa communities. Research Bulletin (Iowa Agriculture and Home Economics Experiment Station) 29:372(1). URL: http://lib.dr.iastate.edu/researchbulletin/vol29/iss372/1.

Ryan, B., Gross, N.C., 1943. The diffusion of hybrid seed corn in two iowa communities. Rural Sociology 8, 15–24.

Stewart, F.A., 1947. A sociometric study of influence in southtown. Sociometry 10, 11. doi:10.2307/2785558.

Valente, T.W., 1996. Social network thresholds in the diffusion of innovations. Social Networks 18, 69–89. doi:10.1016/0378-8733(95)00256-1.

Valente, T.W., 2012. Network interventions. Science 337, 49–53.

Valente, T.W., Vega Yon, G.G., 2020. Diffusion/contagion processes on social networks. Health Education & Behavior 47, 235–248. doi:10.1177/1090198120901497.

Watts, D.J., 2002. A simple model of global cascades on random networks. Proceedings of the National Academy of Sciences 99, 5766–5771. doi:10.1073/pnas.082090499.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. URL: http://ggplot2.org.

Wickham, H., Francois, R., 2016. dplyr: A Grammar of Data Manipulation. URL: https://CRAN.R-project.org/package=dplyr.

Wilensky, U., 1999. NetLogo http://ccl.northwestern.edu/netlogo/. Technical Report. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. URL: http://ccl.northwestern.edu/netlogo/.

van Woudenberg, T.J., Simoski, B., de Mello Araújo, E.F., Bevelander, K.E., Burk, W.J., Smit, C.R., Buijs, L., Klein, M., Buijzen, M., 2019. Identifying influence agents that promote physical activity through the simulation of social network interventions: Agent-based modeling study. Journal of Medical Internet Research 21, e12914. doi:10.2196/12914.

Zhang, J., Shoham, D.A., Tesdahl, E., Gesell, S.B., 2015. Network interventions on physical activity in an afterschool program: An agent-based social network study. American Journal of Public Health 105, S236–S243. doi:10.2105/ajph.2014.302277.

DECLARATIONS for title page

Ethics approval and consent to participate

This study uses anonymised friendship network data collected during the Wellbeing in Schools (WiSe) Survey conducted by The Centre for Evidence and Social Innovation at Queen's University Belfast. For full details of ethics and consent please see (Davison et al., in prep). Briefly, ethical approval has been granted by the School of Social Science, Education and Social Work Ethics Committee, Queen's University Belfast. Participation in the study is by voluntary informed consent, obtained prior to all stages of data collection.

Availability of data and material

WiSe data access is managed by the Centre for Evidence and Social Innovation at Queen's University Belfast. The simulation results supporting the conclusions of this article are available at https://osf.io/kjv4f, together with the NetLogo model used for the simulations.

Competing interests

The authors declare that they have no competing interests.

Funding

The work was undertaken under the auspices of the UKCRC Centre of Excellence for Public Health Research Northern Ireland, which is funded by the British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, the National Institute for Health Research, and the Wellcome Trust (MC CF023241). RH is supported by a NIHR Career Development Fellowship (CDF 2014-07-020) and acknowledges funding support from the HSC Research and Development Division (CDV/5078/14). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements