

RESEARCH

Open Access



# Effectiveness variation in simulated school-based network interventions

Jennifer Badham\* , Frank Kee and Ruth F. Hunter

\*Correspondence:  
research@criticalconnections.com.au  
Centre for Public Health, Queen's  
University Belfast, Institute of Clinical  
Sciences, Block B, Royal Victoria  
Hospital, BT12 6BA Belfast, UK

## Abstract

**Background:** Previous simulation studies have found that starting with high degree seeds leads to faster and more complete diffusion over networks. However, there are few studies and none have used networks that are relevant to a school setting.

**Methods:** We construct 17 networks from friendship nominations in schools and simulate diffusion from a seed group of 15% of the students. That seed group is constructed with seven different approaches (referred to as interventions). The effectiveness of the intervention is measured by the proportion of simulated students reached and the time taken.

**Results:** Seed groups comprising popular students are effective compared to other interventions across a range of measures and simulated contagions. As operationalised, selecting persuasive students is also effective for many simulation scenarios. However, this intervention is not strictly comparable with the others tested.

**Conclusions:** Consistent with previous simulation studies, using popular students as a seed group is a robust approach to optimising network interventions in schools. In addition, researchers should consider supplementing the seed group with influential students.

**Keywords:** Network interventions, Behaviour change, Agent-based modelling, Adolescents

## Background

Behaviour adoption is an important application of diffusion of innovations over networks, for which there is a rich literature (Valente 2005). In public health, the term 'network interventions' specifically refers to using social networks or information about the network to enhance the effectiveness of interventions to encourage healthy behaviour or to discourage unhealthy behaviour (Valente 2012).

Four broad approaches have been described (Valente 2012; Hunter et al. 2017). The first two approaches identify people to promote or participate in the intervention based on their network properties. For the 'Individuals' approach, such identification is based on node level properties, such as degree, that could be expected to increase adoption by other individuals. In contrast, 'Segmentation' selects initial participants based on a shared higher level property such as membership of the same community. 'Induction' interventions encourage greater relevant use of the network by all participants, for example by trying to stimulate discussion. Finally, 'Alteration' interventions are intended to change the network, for example discouraging links with role models of undesirable behaviour.

Focusing on the first approach, a key question for public health researchers is identifying the network properties to be used to select change agents, the people to promote or initially participate in the intervention. The objective is to select the change agents to maximise overall behaviour adoption. In computer science, there is extensive literature developing the 'best' algorithm (however defined) to select the smallest set of starting individuals for a given network and diffusion mechanism to achieve a target diffusion level, or the set of given size that maximises diffusion [originating with Kempe et al. (2003)]. For health behaviour interventions, however, the network and diffusion mechanism may not be known. The research question is therefore identifying network interventions (or change agent selection rules) that are robust; that is, they result in relatively high adoption levels across varied networks and diffusion rules.

Previous research using simulations has found that selection methods that preferentially recruit high degree seed participants lead to greater or faster adoption over several assumed behaviour adoption mechanisms and networks (Aral et al. 2013; Badham et al. 2018). The relative effectiveness of interventions with highly central seeds in these studies is consistent with evidence that indicates that spread is correlated with centrality of seed (de Arruda et al. 2014; Jalili and Perc 2017) under different mechanisms over different networks.

The evidence from real world behaviour change interventions is limited. A recent systematic review found strong evidence of behaviour change with interventions that select popular individuals to lead peer education activities (Hunter et al. 2017). However, in almost all studies reviewed, the control intervention for comparison does not involve peer led interventions; that is, the effect arising from use of popular leaders cannot be separated from the effect of peer led delivery. One empirical study using tickets for access to water purification or vitamin supplements in villages in Honduras specifically compared seed selection methods (Kim et al. 2015). This study found that high degree selection was no more effective than randomly selecting seeds and less effective than asking random individuals to nominate a friend to be the seed.

There are many plausible explanations for the differences between results from simulations and real world intervention trials. Most obviously, the small evidence base may simply be misleading; any general pattern may be obscured by differences in the behaviour adoption mechanisms (Centola 2010; Aral et al. 2013), specific implementation details of the studies or stochastic effects in the diffusion. In addition, it is well established that the structure of the network influences diffusion (Valente 2005), and differences in structure may therefore confound any attempt to compare studies conducted on substantially different networks.

Each of these explanations may be more or less salient for particular intervention applications. Comprehensive simulation studies allow competing explanations to be tested. To investigate these potential explanations, we assess differences in intervention effectiveness across aspects of implementation, specifically ease of diffusion and measures of effectiveness.

We frame our analysis with school based interventions, which are popular for diverse public health behaviours (Jepson et al. 2010; Kriemler et al. 2011; Hale et al. 2014; Das et al. 2016) and also as a setting for network interventions (Hunter et al. 2017). The only network of friendship nominations in existing simulation studies was collected in a prison setting, with the other networks including a large online messaging service, networks

constructed with common algorithms, and observed interactions between adults (Aral et al. 2013; Badham et al. 2018). We therefore use school based friendship networks for the simulation and restrict the seed selection methods to those that are particularly popular in school based network interventions or practical in a school context because they do not require full network analysis. This approach is intended to balance the methodological strengths of simulations with the needs of real-world public health interventions, enhancing the relevance of this study.

## Methods

### Network construction

The networks used in this simulation study were adapted from data collected in the 2016 wave of the Wellbeing in Schools Survey (WiSe) (Davison et al.: Longitudinal overview of adolescent wellbeing: findings from the Wellbeing in Schools Survey, in preparation), when participants were 13–14 years old. All post-primary schools in Northern Ireland ( $N = 181$ ) were invited to participate when WiSe commenced in 2013, and approximately half agreed ( $N = 102$ ). The survey asks students in participating schools about their health and wellbeing, including physical activity, nutrition, drug use and other topics. It is a longitudinal cohort study, surveying one randomly selected class in each participating school every two years.

The social network element of the survey in 2016 asks participants to nominate up to 10 friends in their class, and a checkbox is available for the student to indicate that their friends are not in their class. The respondent is encouraged to fill in their friend's full name. The friendship nomination is recorded in the WiSe dataset using the nominee's study identifier, with a separate code used to indicate that a name was provided but could not be identified for coding. In total, 1603 students participated from 87 schools. Of these, 1369 (85.4%) nominated at least one friend. There were 9144 friend nominations (mean of 6.7 per student), of which 8124 (88.8%) were identifiable.

Of the 87 participating schools, 17 networks satisfied three conditions and were used in this study: (1) the school had at least 20 participating students; (2) at least 80% of the students nominated at least one identifiable friend; and (3) at least 80% of the total school nominations were identifiable. These conditions were set so as to limit missing edges to 36% of the network, at which level the observed network is expected to have similar structural properties as the underlying network (Costenbader and Valente 2003; Kossinets 2006; Smith and Moody 2013), and 80% completion provides an acceptably reliable measure of friendship nominations in a full network (Marks et al. 2012).

These 17 friendship nomination networks were modified for the simulation. Although friendship is a directed relationship, the simulation networks used undirected edges, so adoption influence can be exerted in either direction. In addition, isolates were removed as their adoption status cannot be changed by diffusion mechanisms. The properties of the final networks are summarised at Table 1 and reported individually at Table 2.

Network extraction was undertaken in R (R Core Team 2015, version 3.5.0), with network functionality provided by the 'igraph' package (Csardi and Nepusz 2006, version 1.1.0). The constructed networks were exported into separate gml formatted files for use in the simulation.

**Table 1** Properties of 17 networks used in simulations: Summary

Network property	Mean	SD	Range
Nodes	23.7	3.9	17 to 33 <sup>1</sup>
Edges	97	26	58 to 157
Components	1.2	0.5	1 to 3
Communities	2.9	0.7	2 to 4
Density	0.37	0.09	0.2 to 0.5
Degree: Mean	8.1	1.6	5.0 to 10.7
Degree: Gini	0.19	0.05	0.10 to 0.28
Degree: Assortativity	0.18	0.23	-0.15 to 0.71
Clustering Coefficient: Mean local	0.70	0.09	0.52 to 0.91
Mean shortest path <sup>2</sup>	1.8	0.3	1.4 to 2.7
Diameter <sup>2</sup>	4.1	1.2	2 to 6

<sup>1</sup>While at least 20 contributing students were required for the school to be included, removal of isolates can reduce the corresponding network to fewer than 20 nodes

<sup>2</sup>Calculation of the mean shortest path excluded two networks that were not connected

### Simulation model

An agent-based model was used to simulate behaviour adoption through the network. The model imports one of the networks, selects seed nodes (initial participants) for immediate behaviour adoption, and then simulates diffusion of the behaviour throughout the remainder of the network. The model was developed in NetLogo (Wilensky 1999), a specialist agent-based modelling platform.

Seven different network interventions, or seed selection methods, were available in the model. The operationalisation of the interventions is summarised in Table 3. Note also that these interventions are presented in the same order in all tables and figures.

A common real world network intervention selects opinion leaders to be trained as peer educators, with those leaders selected because of their high degree or other centrality

**Table 2** Properties<sup>1</sup> of 17 networks used in simulations: Individual networks

Network	Nodes	Edges	Comp	Comm	M Degree	G Degree	M Clustering	M Geodesic	Diameter
0	22	84	1	4	7.6	0.22	0.61	1.7	4
1	27	132	1	2	9.8	0.14	0.71	1.8	4
2	21	94	1	3	9.0	0.19	0.63	1.5	3
3	22	115	1	2	10.5	0.13	0.72	1.5	3
4 <sup>2</sup>	24	77	2	3	6.4	0.20	0.82		
5	21	112	1	2	10.7	0.12	0.70	1.4	2
6	22	81	1	3	7.4	0.22	0.74	1.9	5
7	33	157	1	3	9.5	0.19	0.65	2.0	5
8	28	112	1	3	8.0	0.17	0.69	1.9	5
9	17	64	1	2	7.5	0.10	0.91	1.7	3
10	24	89	1	3	7.4	0.21	0.52	1.9	4
11	20	89	1	2	8.9	0.13	0.69	1.6	3
12 <sup>2</sup>	23	58	3	4	5.0	0.26	0.75		
13	26	106	1	4	8.2	0.28	0.64	2.3	6
14	20	61	1	4	6.1	0.19	0.77	2.2	5
15	24	111	1	2	9.3	0.15	0.68	1.7	3
16	29	101	1	3	7.0	0.26	0.68	2.7	6

<sup>1</sup>Comp and Comm refer to the number of components and communities respectively. Where nodal (or nodal pair) properties are summarised, M indicates Mean and G indicates Gini coefficient. Geodesic is the shortest path between a pair of nodes

<sup>2</sup>Networks 4 and 12 are disconnected and geodesics are not calculated

**Table 3** Network interventions simulated

Name	Description <sup>a</sup>
Random Uniform (RU)	Each node has an equal chance of selection
Random by Degree (RD)	Nodes selected with probability proportional to degree
Friend of Random (FR)	A 'nominator' set is randomly selected with equal probability, and each node in that set randomly selects one of their network neighbours to be included in the seed set.
High Degree (HD)	Descending order by degree - number of edges to other nodes
Community Leaders (CL)	Highest degree from each community
Persuasive (P)	Random (uniform) selection, but those selected have greater effect than other nodes during contagion (additional 0.2 transmission probability for simple, or count as two neighbours for complex)
Community (C)	Randomly selected from largest community found by the Louvain optimisation of modularity.

<sup>a</sup>For those interventions where nodes are selected according to rank order, random selection is used for equally ranked nodes if they cannot all be included

measure (Valente 2012; Hunter et al. 2017). The simulated interventions *High Degree* and *Community Leaders* reproduce this selection, with the latter ensuring that the seeds are distributed in different communities (identified by the Louvain algorithm (Blondel et al. 2008)). While popular, these methods require prior collection of network information, and may not always be practical.

Instead, an observer (such as a teacher) could select seeds based on their perception of the participants. *Random by Degree* represents such an observer's attempt at identifying central students, the nodes with higher degree are more likely to be selected. *Persuasive* takes a different approach, selecting those students who are expected to be able to influence others regardless of their centrality. In the simulation, all students are equally likely to be selected, but those selected have a stronger influence in the contagion process with either a higher transmission probability (simple), or double contribution in the calculation of proportion of friends already adopted (complex). Another possible approach where the network is not available is *Friend of Random*, which randomly selects students and then asks each to select one of their friends, which is expected to lead to higher degree seeds by virtue of the friendship paradox (Feld 1991) and has been used in some interventions (such as Kim et al. (2015)).

Finally, the *Community* intervention selects seeds that are within the same community, representing a desire for the seeds to be relatively close in the network so that adopters can provide mutual support (such as Trotter et al. (1996)). Seeds may also be selected randomly and this intervention (*Random Uniform*) also provides a baseline for comparison.

Regardless of the choice of intervention, the number of seeds selected represented 15% (rounded up) of the network. This reflects the common recruitment target to achieve critical mass in public health interventions (Kelly and Stevenson 1995). Previous research has found that the initial adoption level has limited impact on relative effectiveness of interventions (Badham et al. 2018). The selected nodes were assigned as having adopted the behaviour at the start of the simulation, with all other nodes as not adopted.

Two diffusion mechanisms were simulated. Simple contagion (Centola and Macy 2007) has all nodes that have already adopted 'transmit' the behaviour to their network neighbours, who then adopt with some probability. The transmission probability is a model parameter, with value of 0.4, 0.7 or 1. These were chosen to span the informative range;

from exploratory simulations, the differences between interventions were obscured by randomness in the diffusion at lower probability values. For complex contagion (Valente 1996; Centola and Macy 2007), each node that has not already adopted calculates the proportion of its network neighbours who have adopted, and then adopts if that proportion meets or exceeds a specified threshold (a model parameter, with value between 0.2 and 0.7 in increments of 0.1).

### Experimental design

Overall, there were 1,071 combinations of model settings available in a full factorial design (17 networks by 7 interventions by 9 diffusion mechanisms and parameter values). The experimental design is summarised at Table 4, requiring 317,220 simulations overall. These simulations were managed with BehaviorSpace, the batch simulation tool within NetLogo. Multiple simulation runs are required where there is randomness either in the selection of seeds or the behaviour adoption mechanism. The multiple runs allows an average outcome to be calculated, so that effects can be more appropriately compared.

Two of the interventions are deterministic, so they will select the same seeds given the same network (except as necessary to break ties). The other five interventions are stochastic. For the transmission process, complex contagion is deterministic as is simple contagion with transmission probability of 1. However, to simplify the set up, the simple contagion simulations with probability of 1 were included in the same set of experiments as the other (stochastic) simple contagion simulations.

The deterministic simulation sets were run 5 times, to allow for some variation if ranks are tied. Those with one source of randomness (either simple contagion, or one of the stochastic interventions) were run 100 times. Those simulations with both potential sources of randomness were run 1000 times.

The simulation stops when no further diffusion can occur. For each time step in the simulation, the key measure reported is the number of nodes who have adopted the behaviour.

Four measures were derived from the simulation output for each run. Two concern the proportion of the network adopted after one and two time steps ('1-hop reach' and '2-hop reach' respectively), which estimate the potential impact of the intervention assuming that later adopters are less likely to diffuse the intervention further. The other two concern the status of the simulation when completed, the proportion of nodes adopted ('penetration') and the number of time steps to achieve that level of penetration ('duration').

**Table 4** Experimental design: simulation parameters and number of runs

Parameter	Values
Network	17 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 proportions of 0.2 to 0.7 by 0.1
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

Each measure was calculated as the mean over all simulations (5, 100 or 1000) with the same simulation parameters: intervention, network, contagion type and transmission probability or threshold.

For complex contagion simulations only, it is possible that no contagion occurs and the final adoption level is simply the 15% of individuals initially selected. Therefore, additional effectiveness measures were calculated: the proportion of simulations where there was at least one secondary adoption (referred to as a cascade), and the proportion of the network adopted ('1-hop reach', '2-hop reach' and 'penetration') given that secondary adoption occurred.

The results were analysed using R (R Core Team 2015), particularly the dplyr package (Wickham and Francois 2016, v 0.5.0).

## Results

The question of interest is the relative effectiveness of the network interventions over the various measures described above: proportion of simulations where secondary adoptions occurred, adoption level achieved, and number of time steps to achieve final adoption levels. Further, where there is inconsistency across different sets of simulations for a specific measure, it is important to understand whether this variation is more attributable to the measure of effectiveness used or associated with specific simulation parameters such as the probability of transmission or threshold. Simple and complex contagion simulations are reported separately.

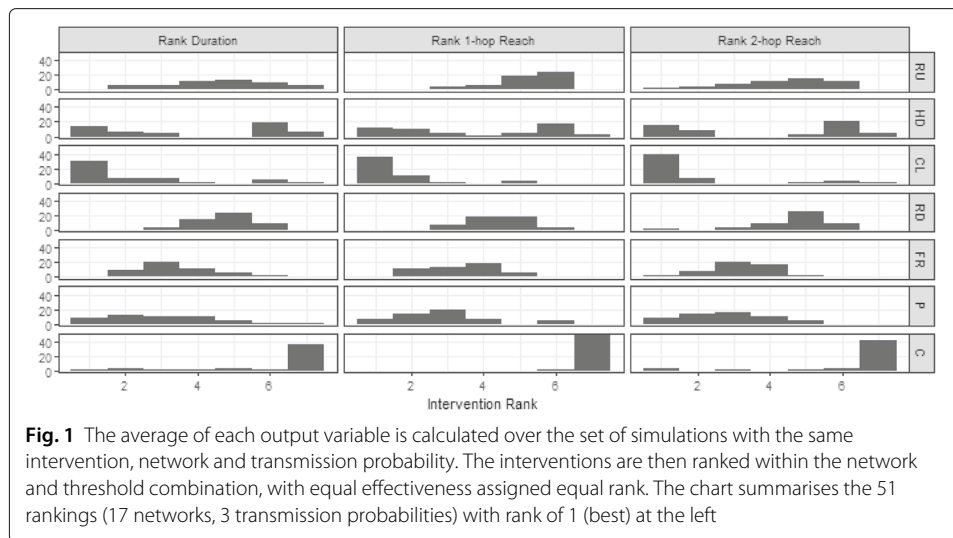
### Simple contagion: probabilistic

For simple contagion, we found limited impact of transmission probability on effectiveness measures except for *Persuasive*, with similar rank patterns for all other interventions (Additional file 1: Figure S3). For simple contagion, the *Persuasive* intervention selects the seeds randomly but assigns them 0.2 higher probability of transmission; for example, in the 0.4 transmission probability scenarios, most nodes have probability of transmission to their neighbours of 0.4 at each time step, but seeds have probability 0.6. At baseline transmission probability of 1, this intervention has no effect and is equivalent to *Random Uniform*. At lower baseline probabilities, the additional 0.2 has a larger effect and the ranking of *Persuasive* improves to second or third, shifting the other interventions to a lower rank.

The rank pattern for each intervention is similar over different effectiveness measures (see Fig. 1, also Additional file 1: Figure S4). The most effective intervention is *Community Leaders* or *High Degree* for all three measures in almost all simulation sets. On the other hand, *High Degree* is also ranked poorly in many simulations. Apart from the low ranked *High Degree*, the order of the remaining interventions over all simulations is reasonably consistent as: *Persuasive*, *Friend of Random*, *Random by Degree*, *Random Uniform*, with *Community* almost always the least effective.

While not apparent in the ranking visualisation, there is a ceiling effect for some of the effectiveness measures. For example, of the 105 simulation sets with transmission probability of 0.7 on single component networks (15 networks and 7 interventions), 70 simulation sets have a 2-hop reach of at least 0.95. That is, almost all the interventions lead to adoption across the entire network in two time steps for almost all networks. In such a situation, rankings may distinguish between very small differences in the proportion





of simulations that achieved complete adoption rather than a meaningful difference in intervention effectiveness. In contrast, 1-hop reach displays a greater range of proportion of network adopted, with only six simulation sets with reach  $\geq 0.9$  and nine sets with reach  $\leq 0.6$ . The observation of similar ranking patterns across measures (Fig. 1) suggests that results may be generalisable despite this ceiling effect.

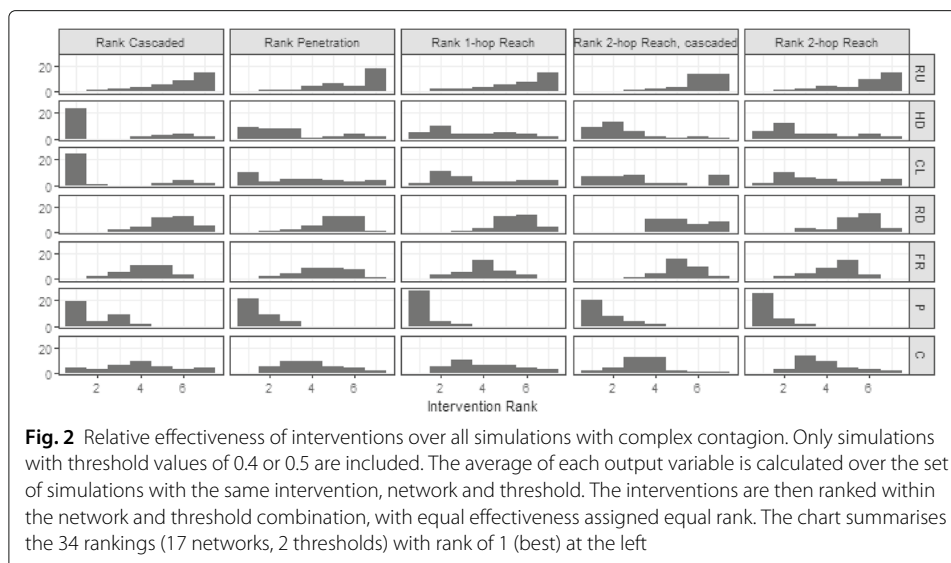
#### Complex contagion: threshold

For complex contagion, the simulations lead to unrealistically high adoption at low thresholds (full adoption in 93% of simulations for 0.2 and 70% for 0.3); it is too easy for a small proportion of friends who have already adopted to trigger a new adoption. At the other extreme, with a threshold of 0.7, only 9% of simulations have any adoptions other than the seeds, with 21% at a threshold of 0.6. Except as stated, further analysis is therefore restricted to the simulations with thresholds of 0.4 or 0.5, for which there is the greatest scope for differentiation between simulations with different parameters.

The pattern of intervention rankings in complex contagion simulations is consistent over effectiveness measures but varies by threshold more substantially than the equivalent analysis over transmission probability for simple contagion (see Fig. 2 and Additional file 1: Figure S5). Broadly, the most effective intervention is *Persuasive* for all measures in almost all simulation sets. Unlike the case of simple contagion, the implementation of *Persuasive* confers an advantage at any threshold, as the selected individuals always contribute double to the calculation of proportion already adopted. This is followed by *High Degree* and *Community Leaders*. The next most effective intervention is *Community*, followed by *Friend of Random*. Finally, *Random Uniform* is almost always the least effective.

As for simple contagion, *High Degree* shows the greatest variation in relative effectiveness. This intervention is relatively effective at threshold of 0.4, with mixed results for threshold of 0.5. For example, assessed with 2-hop Reach at 0.5 threshold, *High Degree* performs well (rank 1 or 2) over seven of the 17 networks and poorly (rank 6 or 7) over five. Unlike simple contagion, *Community Leaders* shows a similar mixed pattern as *High Degree*.





The most dissimilar result between the two types of contagion is for the *Community* intervention, which is relatively ineffective in simple contagion but is generally effective for complex contagion. That is, drawing seeds from within a group of friends is useful where contagion is linked to the proportion of friends already adopted because those seeds are more likely to have friends in common.

#### Comparing *High Degree* to *Random Uniform*

The *High Degree* intervention is particularly interesting. It is the intervention expected to be relatively effective based on successful interventions (Valente 2012; Hunter et al. 2017) and previous simulation studies (Aral et al. 2013; Badham et al. 2018). However, it is also the intervention that was relatively ineffective in the only health behaviour trial that directly compares network seed selection methods (Kim et al. 2015), and displays considerable inconsistency in simulation results in this study. We therefore report the simulation results in more detail for *High Degree*, directly comparing it to *Random Uniform* selection.

There are 153 pairs of results to compare for simple contagion: 17 networks, 3 transmission probabilities, and 3 effectiveness measures (1-hop reach, 2-hop reach, and duration). Of these, *High Degree* is more effective in 80 pairs, *Random Uniform* is more effective in 71 pairs, with equal effectiveness in the other 2 pairs.

Part of this difference in the relative effectiveness of the two interventions is a network effect. Of the 17 networks, *High Degree* is more effective for all three measures and all three transmission probabilities over seven networks (0, 2, 4, 5, 8, 9, 11), accounting for 63 of the 153 simulation pairs. There are a further seven networks (1, 6, 7, 13, 14, 15, 16) where *Random Uniform* is more effective for both duration and 2-hop reach with all three transmission probabilities (42 simulation pairs). However, *Random Uniform* is consistently effective for 1-hop reach on only three of these networks. An examination of the networks shows no clear difference in their properties (see Table 2) where each intervention is more effective, except for a tendency for better results for *High Degree* on networks with a larger diameter.

There are additional measures of effectiveness for complex contagion because of the need to take into account whether any secondary adoptions occurred. To provide more

evidence about such cascading, this comparison between the *High Degree* and *Random Uniform* includes simulations with a threshold of 0.6, as well as the 0.4 and 0.5 analysed previously. As a result, there are 357 pairs of results to compare for complex contagion: 17 networks, 3 thresholds, and 7 effectiveness measures (proportion cascading, and 1-hop reach, 2-hop reach and penetration over all simulations and just for those where cascades occurred).

From Table 5, it is clear that the *High Degree* intervention is relatively effective at the lower threshold values, where it is easier to generate adoption cascades. In contrast, this intervention is relatively ineffective with a threshold of 0.6 for those measures where all simulations are included.

The proportion cascaded measure indicates the source of this reversal. With higher thresholds, the *Random Uniform* intervention is more successful at triggering at least some secondary adoptions. Restricting only to those simulations with such secondary adoptions, *High Degree* seeds are able to reach a greater proportion of the network with one or two time steps for 14 of the 17 networks. On the other hand, for 16 networks, *Random Uniform* seeds eventually achieve a greater proportion of the network adopted, even restricting to those simulations where at least one secondary adoption occurs.

These results indicate different patterns of secondary adoption. *High Degree* tends toward an ‘all or nothing’ pattern where lower success in triggering any adoptions is combined with larger numbers of such adoptions when they are triggered. In contrast, *Random Uniform* tends toward a ‘slow and steady’ pattern where there are more and longer chains of small numbers of additional adoptions.

See also Additional file 1: Figures S6, S7 and S8 for selected intervention rank sequences that display the variation in effectiveness of *High Degree* and *Community Leaders*. The remaining interventions maintain their relative ranks as these two interventions vary position.

## Discussion

Our longer term goal is to develop guidance concerning seed selection for health behaviour interventions in a school setting. This study explores potential explanations for inconsistencies in the literature about the role of degree. Previous simulation studies (Aral et al. 2013; Badham et al. 2018) found that selecting highly central starting nodes is generally more effective (faster or more complete adoption) than uniform random selection, but provided limited guidance on the situations in which such selection may

**Table 5** Number of networks over which *High Degree* (HD) or *Random Uniform* (RU) intervention is more effective, complex contagion

Measure <sup>1</sup>	Threshold 0.4		Threshold 0.5		Threshold 0.6	
	HD	RU	HD	RU	HD	RU
1-hop reach	15	2	11	6	3	14
2-hop reach	15	2	11	6	2	15
Penetration	14	3	12	5	2	15
Proportion cascaded	15	2	11	6	3	14
1-hop reach, cascaded	15	2	12	5	14	3
2-hop reach, cascaded	16	1	16	1	14	3
Penetration, cascaded	13	4	11	6	1	16

<sup>1</sup>Each effectiveness measure is calculated as the average over 5 simulations for the *High Degree* intervention and 100 simulations for the *Random Uniform* intervention

fail. Within the context of school friendship networks, this study supports those results and, further, demonstrates that the increased effectiveness arises whether those seeds are recruited across the network (*High Degree*) or distributed between communities (*Community Leaders*).

However, there is a non-trivial risk that fewer people would adopt than would have adopted with randomly selected seeds. For simple contagion, this risk is independent of the transmission probability and the measure of effectiveness. While it appears to be related to the network structure, particularly path lengths, further work is required to understand the combination of properties that contributes to lower adoption with high degree seeds.

For complex contagion, the risk of reduced adoption is lower but also has a more complicated pattern. For those simulations where contagion is relatively easy (low threshold), the *High Degree* intervention is consistently effective over different measures and networks. However, as the required proportion of peers increases, high degree seeds are more likely than randomly selected seeds to fail to trigger any secondary adoptions. If secondary adoptions are triggered, using high degree seeds results in a larger number of these secondary adoptions initially but not over the whole of the simulation. *High Degree* is therefore relatively ineffective for those simulations where contagion is difficult to achieve. In such a situation of difficult diffusion, however, any network approach is unlikely to be the most appropriate intervention design. There is an additional network effect not clearly related to a specific structural property.

As operationalised, the *Persuasive* intervention is very effective for both contagion types. This intervention is not strictly comparable to the others as the contagion mechanism is altered by the operationalisation. Further, the arbitrary operationalisation means that the size of the simulated effect provides no evidence about the effect in the real world, or how to identify suitable individuals. Nevertheless, the simulations suggest that the personal characteristics of initial seeds may be more important than their network positions. The ASSIST intervention (Campbell et al. 2008) recruited nominated influential students to promote smoking prevention messages within their social networks. The success of that intervention provides evidence that the effectiveness of the simulated *Persuasive* intervention can be translated to a real world school setting. An additional advantage of this approach is that nomination of observable behaviour such as influence is more robust to missing network data than private relationships such as friendship (Marks et al. 2012).

Finally, *Community* is relatively successful for complex contagion simulations, with similar patterns as for *High Degree* and *Community Leaders*. However, it is consistently ineffective for simple contagion. As real world interventions are likely to include aspects of both simple contagion (such as information provision) and complex contagion (such as social norms) and there is disadvantage associated with the *Community* intervention, this study does not support its use.

A key limitation of this simulation study is its use of idealised network diffusion mechanisms. Each implements behaviour adoption as the outcome of an entirely social contagion process; exposure or awareness via social contacts (simple), or as compliance with social norms (complex). While the fundamental mechanisms of behaviour change are not well understood (Michie et al. 2014), it is clear that network diffusion is only one of several factors and may have only a small influence. Other factors include individual attributes such as attitude and environmental factors such as adoption costs.

Several approaches have been taken to represent the contribution of multiple factors in agent-based models of health behaviour change (Yang 2019). For example, utility functions combine the costs and benefits of each factor and the choice with the highest utility value is selected by the agent. Similarly, decision trees can represent a set of factor specific conditions that are considered sequentially to reach a decision. The choice of decision model affects how social influence is included. Further, the concept of social influence is often underspecified and ignores aspects like personality, affect and variable social cognition, characteristics that may have a bearing on behavioural transmission. Designing the most effective interventions will need to take such mechanistic heterogeneity into account.

A further limitation concerns the symmetrisation of the WiSe networks, with the directed friendship relationship treated as undirected for the simulation. This approach implicitly assumes that mechanisms of action for the behaviour change represented by the simulation can occur in both directions, such as emulating behaviour in someone the individual considers a friend (nominee to nominator), or enthusiastically promoting a behaviour to others (nominator to nominee). In practice, diffusion may occur primarily or only in one direction. Interventions that use high degree opinion leaders often select on the basis of in-degree only but also use relationships that are more explicitly influential than simply friendship (at least in public health (Valente 2012; Hunter et al. 2017)).

## Conclusions

Ideally, detailed information about networks and the relevant diffusion mechanisms would allow specific simulations to optimise the intervention for each specific situation. In the absence of such information, we would recommend selecting popular students (high degree) to act as seeds, especially if there is reason to believe that the proportion of friends already adopted has a role in behaviour change.

While the friendship network provides support for the intervention diffusion, the effectiveness of *Persuasion* suggests that additional network questions are warranted. Surveys intended to provide information for health behaviour interventions could include nomination of an influential person in addition to friends. This would allow the intervention to include both popular and influential students in the seed group.

More work is clearly needed to refine this advice, with identification of the relevant network properties for differential effectiveness constituting the most substantial knowledge gap in the idealised setting of network diffusion. For translation to real world network interventions, greater understanding of behaviour change mechanisms is required to assess the potential effectiveness gain available through optimal selection.

## Additional file

**Additional file 1:** Supplementary **Figures S3 to S8** are available as a download. (PDF 345 kb)

## Abbreviations

C: *Community* intervention; CL: *Community leaders* intervention; FR: *Friend of random* intervention; HD: *High degree* intervention; P: *Persuasive* intervention; RD: *Random by degree* intervention; RU: *Random uniform* intervention

## Acknowledgements

The authors would like to thank the reviewers for their helpful comments.

### Authors' contributions

This study was designed by JB and RH. JB developed the model, conducted the analysis and prepared the first draft of the paper. RH developed the social networks element of the WiSe survey, together with the WiSe research team. All authors contributed to the final version and have approved its submission.

### 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.

### Availability of data and materials

WiSe data access is managed by the Centre for Evidence and Social Innovation at Queen's University Belfast.

Supplementary materials for this project are archived at <https://osf.io/kjv4f>. These comprise compressed files for each of:

- Model code [Model.zip] - requires NetLogo to run; and
- Simulation results [Simulation.zip] - csv files with a data dictionary.

NetLogo version 6 (Wilensky 1999) must be installed to run the model. This software is available as a free download for Windows, Mac and Linux.

The specific model can be downloaded from the archive as a compressed (zip) file with the network data. While the model can be saved with any path and name, the folder with the network data must be saved in the same location and names cannot be altered. This is to allow the NetLogo model to import the network structure files as required.

### 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.: Longitudinal overview of adolescent wellbeing: findings from the Wellbeing in Schools Survey, in preparation). 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.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 21 March 2019 Accepted: 2 July 2019

Published online: 23 September 2019

### References

- Aral S, Muchnik L, Sundararajan A (2013) Engineering social contagions: Optimal network seeding in the presence of homophily. *Netw Sci* 1(02):125–153. <https://doi.org/10.1017/nws.2013.6>
- Badham J, Kee F, Hunter RF (2018) Simulating network intervention strategies: Implications for adoption of behaviour. *Netw Sci* 6(02):265–280. <https://doi.org/10.1017/nws.2018.4>
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008(10):10008
- Campbell R, Starkey F, Holliday J, Audrey S, Bloor M, Parry-Langdon N, Hughes R, Moore L (2008) An informal school-based peer-led intervention for smoking prevention in adolescence (ASSIST): a cluster randomised trial. *Lancet* 371(9624):1595–1602. [https://doi.org/10.1016/s0140-6736\(08\)60692-3](https://doi.org/10.1016/s0140-6736(08)60692-3)
- Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329(5996):1194–1197
- Centola D, Macy M (2007) Complex contagion and the weakness of long ties. *Am J Sociol* 113(3):702–734
- Costenbader E, Valente TW (2003) The stability of centrality measures when networks are sampled. *Soc Netw* 25(4):283–307. [https://doi.org/10.1016/s0378-8733\(03\)00012-1](https://doi.org/10.1016/s0378-8733(03)00012-1)
- Csardi G, Nepusz T (2006) The igraph software package for complex network research. *InterJournal Complex Systems*:1695
- Das JK, Salam RA, Arshad A, Finkelstein Y, Bhutta ZA (2016) Interventions for adolescent substance abuse: An overview of systematic reviews. *J Adolesc Health* 59(4):61–75. <https://doi.org/10.1016/j.jadohealth.2016.06.021>
- de Arruda GF, Barbieri AL, Rodriguez PM, Rodrigues FA, Moreno Y, da Fontoura Costa L (2014) Role of centrality for the identification of influential spreaders in complex networks. *Phys Rev E* 90(3). <https://doi.org/10.1103/physreve.90.032812>
- Feld SL (1991) Why your friends have more friends than you do. *Am J Sociol* 96(6):1464–1477. <https://doi.org/10.1086/229693>
- Hale DR, Fitzgerald-Yau N, Viner RM (2014) A systematic review of effective interventions for reducing multiple health risk behaviors in adolescence. *Am J Public Health* 104(5):19–41. <https://doi.org/10.2105/ajph.2014.301874>
- Hunter RF, de la Haye K, Badham J, Valente T, Clarke M, Kee F (2017) Social network interventions for health behaviour change: a systematic review. *Lancet* 390:47. [https://doi.org/10.1016/s0140-6736\(17\)32982-3](https://doi.org/10.1016/s0140-6736(17)32982-3)
- Jalili M, Perc M (2017) Information cascades in complex networks. *J Complex Netw.* <https://doi.org/10.1093/comnet/cnx019>

- Jepson RG, Harris FM, Platt S, Tannahill C (2010) The effectiveness of interventions to change six health behaviours: a review of reviews. *BMC Public Health* 10(1). <https://doi.org/10.1186/1471-2458-10-538>
- Kelly J, Stevenson L (1995) Opinion Leader HIV Prevention Training Manual. Center for AIDS Intervention Research, Medical College of Wisconsin, Milwaukee
- Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '03. ACM Press. <https://doi.org/10.1145/956750.956769>
- Kim DA, Hwong AR, Stafford D, Hughes DA, O'Malley AJ, Fowler JH, Christakis NA (2015) Social network targeting to maximise population behaviour change: A cluster randomised controlled trial. *Lancet* 386(9989):145–153. [https://doi.org/10.1016/s0140-6736\(15\)60095-2](https://doi.org/10.1016/s0140-6736(15)60095-2)
- Kossinets G (2006) Effects of missing data in social networks. *Soc Netw* 28(3):247–268. <https://doi.org/10.1016/j.socnet.2005.07.002>
- Kriemler S, Meyer U, Martin E, van Sluijs EMF, Andersen LB, Martin BW (2011) Effect of school-based interventions on physical activity and fitness in children and adolescents: a review of reviews and systematic update. *Br J Sports Med* 45(11):923–930. <https://doi.org/10.1136/bjsports-2011-090186>. <https://bjsm.bmj.com/content/45/11/923.full.pdf>
- Marks PEL, Babcock B, Cillessen AHN, Crick NR (2012) The effects of participation rate on the internal reliability of peer nomination measures. *Soc Dev* 22(3):609–622. <https://doi.org/10.1111/j.1467-9507.2012.00661.x>
- Michie S, West R, Campbell R, Brown J, Gainforth H (2014) ABC of Behaviour Change Theories. Silverback Publishing, Great Britain
- R Core Team (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>
- Smith JA, Moody J (2013) Structural effects of network sampling coverage I: Nodes missing at random. *Soc Netw* 35(4):652–668. <https://doi.org/10.1016/j.socnet.2013.09.003>
- Trotter RT, Bowen AM, Baldwin JA, Price LJ (1996) The efficacy of network-based HIV/AIDS risk reduction programs in midsized towns in the united states. *J Drug Issues* 26(3):591–605. <https://doi.org/10.1177/002204269602600305>
- Valente TW (1996) Social network thresholds in the diffusion of innovations. *Soc Netw* 18(1):69–89. [https://doi.org/10.1016/0378-8733\(95\)00256-1](https://doi.org/10.1016/0378-8733(95)00256-1)
- Valente TW (2005) Network models and methods for studying the diffusion of innovations. In: Carrington PJ, Scott J, Wasserman S (eds). *Models and Methods in Social Network Analysis*. Cambridge University Press. pp 98–116. <https://doi.org/10.1017/cbo9780511811395.006>
- Valente TW (2012) Network interventions. *Science* 337(6090):49–53
- Wickham H, Francois R (2016) dplyr: A Grammar of Data Manipulation. <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. <http://ccl.northwestern.edu/netlogo/>
- Yang Y (2019) A narrative review of the use of agent-based modeling in health behavior and behavior intervention. *Transl Behav Med*. <https://doi.org/10.1093/tbm/iby132>

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---