

# Network properties of the school system in England: Implications for the effect of school reopening strategies on transmission of SARS-COV-2

James D Munday\*, Katharine Sherratt, Sophie Meakin, Joel Hellewell, Sam Abbott, Nikos I Bosse, CMMID COVID-19 Working Group, W. John Edmunds, Albert Jan van Hoek, Sebastian Funk  
*Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine*  
\*Correspondence: james.munday@lshtm.ac.uk

## Abstract

### *Background*

Closing schools has been a frequent non-pharmaceutical intervention against the spread of COVID-19 across the world including the UK. As governments begin to relax restrictions on public life there is a need to understand the potential impact that reopening schools will have on transmission.

### *Methods*

We constructed a network of transmission between schools and households in England using data provided by the HM Government, Department for Education (DfE). We evaluated the network to assess the potential for transmission between adjacent schools and for long range transmission across the network under different reopening scenarios and values of the within-school reproduction number.

### *Results*

We found that reopening only Reception, Year 1 and Year 6 resulted in the lowest risk of transmission between schools, with outbreaks within a single school unlikely to result in outbreaks in neighbouring schools. Reopening either or both of year 10 and 12 in addition resulted in similar risk of transmission to reopening all primary school years resulting in some local transmission but with the majority of schools remaining in small components in the network. Reopening all secondary years resulted in large scale transmission between schools. Increased within-school reproduction number increased rate of transmission between schools.

### *Conclusions*

Reopening secondary school years is likely to have a greater impact on transmission than reopening primary school years. Maintaining a low within-school reproduction number is essential in reducing risk of large outbreaks amongst school aged children and their families.

### *Limitations*

- The network only captures transmission in schools and households between children within the school system
- The data from which the network is constructed includes only state funded schools in England with children coded as school years Reception to Year 13
- The quantification of the probability of transmission between schools assumed that each school outbreak reached its “natural” final size
- Transmission within schools is assumed to be homogeneous and the reproduction number is assumed to be the same in all schools
- Transmission between children is assumed to be common with a reproduction number in schools of greater than 1.

## Introduction

School closures are one of many non-pharmaceutical interventions that can be applied during outbreaks, and can be highly effective in reducing transmission.<sup>1</sup> However, there are significant societal costs associated with closing large numbers of schools, such as limiting children's access to education and requiring caregivers to stay at home (impacting on household income and on economic activity).<sup>2</sup> In addition, closures can affect single parent and deprived households most.<sup>3</sup> It is therefore important for policy makers to plan school closures well to minimise the economic and social impact, whilst remaining effective in curbing the outbreak.

School closures were introduced as a central component of the COVID-19 response in many countries around the world<sup>4</sup>. The UK closed all schools on 23<sup>rd</sup> March 2020 and they remain closed to this day. The government is now considering various strategies for reopening schools in the coming weeks and months.

Although reported cases of COVID-19 are low amongst school-aged children, the role of children in transmitting COVID-19 is unclear.<sup>5,6,7,8</sup> There is a risk that reopening schools will increase transmission within the community and therefore lead to a resurgence of the virus in the population. The reopening of schools could increase the risk in transmission in two ways: schools may act as a route for transmission between households, and households with multiple school-aged children may act as a route for transmission between schools. The combination of these factors has the potential to create large networks of connected contacts which expose a large proportion of households to additional risk of infection.

We use a network model of school and household transmission to explicitly evaluate the potential for various school reopening programs to increase risk of widespread transmission within the community by school to school transmission, through households.

---

<sup>1</sup> Wu et al. School Closure and Mitigation of Pandemic (H1N1) 2009, Hong Kong. (2020). *Emerging Infectious Diseases*.

<sup>2</sup> Viner, R.M., Russell, S.J., Croker, H., Packer, J., Ward, J., Stansfield, C., Mytton, O., Bonell, C. and Booy, R., 2020. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. *The Lancet Child & Adolescent Health*.

<sup>3</sup> Van Lancker W, Parolin Z. COVID-19, school closures, and child poverty: a social crisis in the making. *Lancet Public Health*. 2020;5(5):e243-e244. doi:10.1016/S2468-2667(20)30084-0

<sup>4</sup> Tracking Public Health and Social Measures. World Health Organisation.

<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/phsm>. Accessed 17 May 2020.

<sup>5</sup> Dong Y, Mo X, Hu Y, Qi X, Jiang F, Jiang Z, et al. Epidemiological Characteristics of 2143 Pediatric Patients With 2019 Coronavirus Disease in China. *Pediatrics* 2020:e20200702. doi:10.1542/peds.2020-0702

<sup>6</sup> Bi Q, Wu Y, Mei S, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study, *Lancet Infectious Disease*, April 27th 2020, [https://doi.org/10.1016/S1473-3099\(20\)30287-5](https://doi.org/10.1016/S1473-3099(20)30287-5)

<sup>7</sup> Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic Population. *N Engl J Med*, Published April 14th 2020, doi:10.1056/NEJMoa2006100.

<sup>8</sup> Nicholas G Davies, Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, CMMID COVID-19 working group, Rosalind M Eggo. Age-dependent effects in the transmission and control of COVID-19 epidemics. medRxiv 2020.03.24.20043018; doi: <https://doi.org/10.1101/2020.03.24.20043018>

## Methods

### *Data*

Individual level anonymised pupil data has been provided by the Department for Education. The data includes an entry for each pupil for each institution they are in attendance at, Unique Reference Number (URN) for the school, school postcode, pupil's postcode and pupil's address. We combined the student's postcode and address to assign a code for each group of pupils that were found to live at the same address, assuming these are "households". Using this code we were able to calculate the number of unique contacts between each pair of schools. For each pupil we have included only institutions coded as the pupil's main school.

### *Scenarios*

We consider 6 scenarios, where particular year groups return to school.

1. Reception, Year 1 and Year 6.
2. As scenario 1 with Year 10
3. As scenario 1 with Year 12
4. As scenario 1 with Year 10 and Year 12
5. All primary school years (Reception – Year 6)
6. All secondary school years (Years 7 – 13)

For each scenario we generated a network with only pupils within the years specified to ensure only the correct household links were included in the network.

### *A network of schools*

We constructed a network of all state-funded schools in England, as linked through households. Each edge on the school-level contact network is weighted by the number of unique contacts between schools that occur through shared households. For example, if in a given household, 2 children attend school  $A$  and 2 children attend school  $B$ , this corresponds to 4 unique contacts between school  $A$  and school  $B$ . The total number of unique contacts between schools  $A$  and  $B$ , denoted by  $C_{AB}$ , is the sum of unique contacts over all households. (Figure 1)

From this network, we created a transmission probability network (Figure 2) where we estimated the probability of transmission between schools  $i$  and  $j$  as:

$$P_{trans, ij} = 1 - \prod (1 - P_{ob} P_j^I q)^{C_{ij}}$$

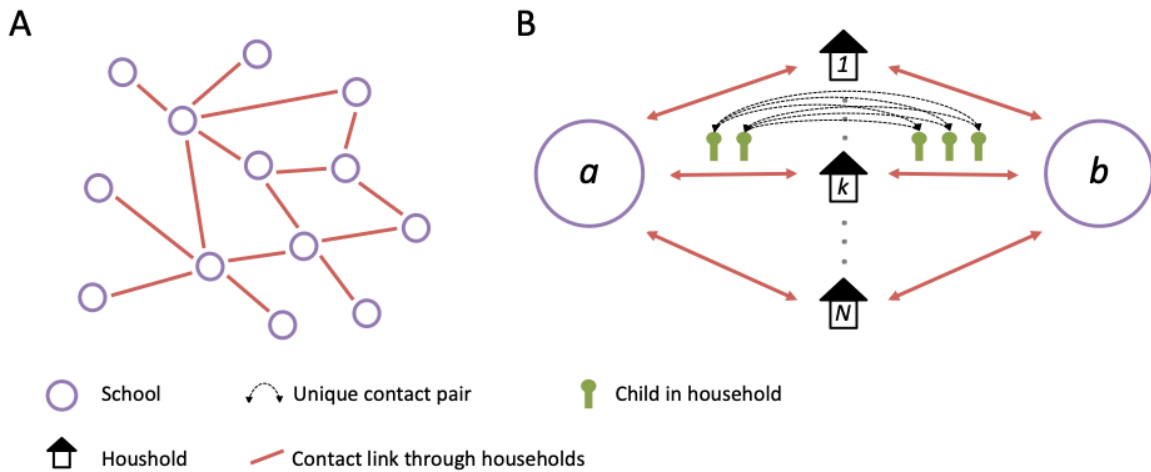
where probability of an outbreak  $P_{ob}$  in any given school is,

$$P_{ob} = 1 - 1/R,$$

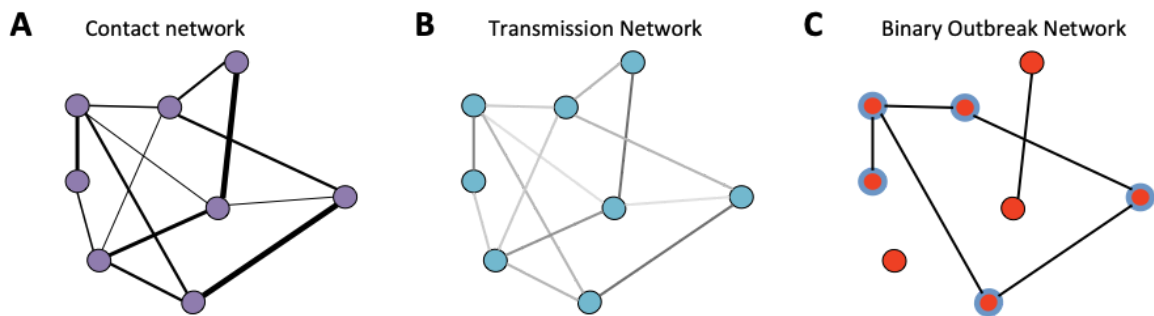
and the probability of a student in school  $j$  being infected  $P_j^I$  is,

$$P_j^I = R_\infty = 1 - e^{-R R_\infty},$$

where  $R$  is the reproduction number within school,  $R_\infty$  is the final outbreak size as a proportion of all students in the school and  $q$  is the per-contact probability of transmission between children in the same household. The value of  $q$  has been fixed at  $0.15^9$  for the purposes of this analysis.



**Figure 1.** A network of schools linked by households. A) A network of schools constructed such that schools are connected when contact is made between pupils of different schools within a household. B) The strength of contact between schools is quantified by calculating the number of unique contact pairs (one child in each school). The number of pairs per household is the product of the number of children who attend school  $a$  and the number of children who attend school  $b$ . The total number of unique pairs is the sum of unique pairs in each household with children attending both school  $a$  and  $b$ .



**Figure 2** A) A schematic of a contact network, the width of the edges shows the relative probability of transmission between schools B) A schematic of a transmission probability network calculated from the contact network, the shading of the edges shows the relative probability of transmission between schools. C) A schematic of a realisation of a binary outbreak network (sampled from A), where edges are weighted 1 with probability given by the equivalent edge in the transmission network, or 0 otherwise. Blue highlighted nodes show those in the largest connected component. In each network nodes show the location of schools.

<sup>9</sup> Wee Chian Koh, Lin Naing, Muhammad Ali Rosledzana, Mohammad Fathi Alikhan, Liling Chaw, Matthew Griffith, Roberta Pastore, Justin Wong, What do we know about SARS-CoV-2 transmission? A systematic review and meta-analysis of the secondary attack rate, serial interval, and asymptomatic infection, medRxiv 2020.05.21.20108746; doi: <https://doi.org/10.1101/2020.05.21.20108746>

## ***Evaluating the network***

### *Degree distributions*

To summarise how the potential of transmitting to adjacent schools in the network varies with R and the reopening scenario we calculated the weighted degree distribution of the transmission network (the distribution of the expected number of schools infected through households by each school) for each scenario.

### *Connected components of binary transmission networks*

To summarise the potential spread of the virus across the network of schools, we sample instances of binary transmission networks, where transmission between each pair of schools either occurs (edge weight of 1) or does not occur (edge weight 0) (Figure 2). The possible spread of an outbreak from each school is defined as the connected component of each school. Because schools vary in size considerably, with large differences between secondary and primary schools we also calculated the number of households with children attending schools within each connected component in the network. To establish the number of households represented in each connected component we calculated the number of unique households with children attending the schools in each component in the appropriate years for each scenario from the pupil records data.

To summarise the potential for the virus to spread under certain policy conditions and with different transmission characteristics within schools, we present the component size distribution for realisations of the binary transmission networks and the distribution of the number of households associated with each connected component..

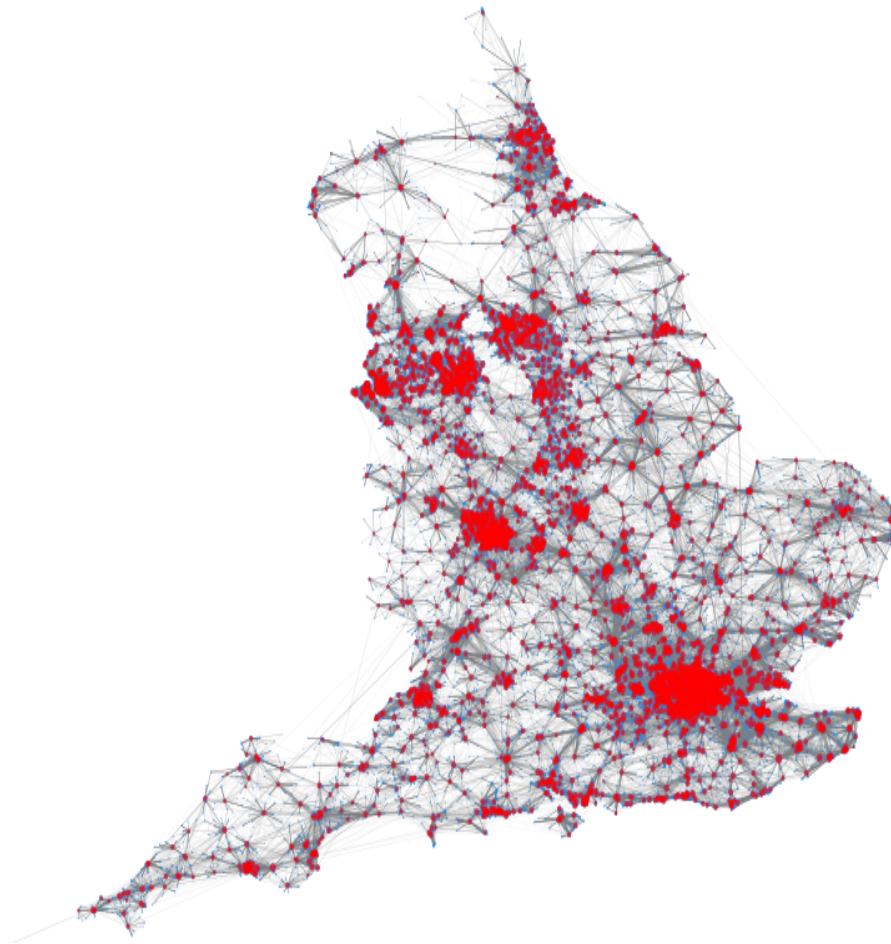
## **Results**

### ***Network summary***

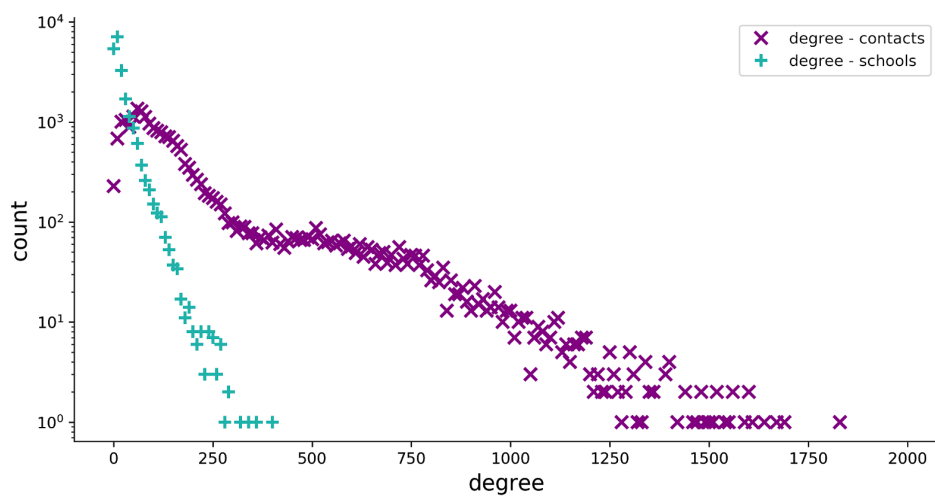
The full contact network (Figure 3) included 21,608 individual schools containing 4.6 million primary school children and 3.4 million secondary school children living at 4.9 million addresses.

The mean unweighted degree of the network (number of schools connected) was 26 with a maximum of 401. The mean number of contact pairs to all other schools was 189 with a maximum of 2088. (Figure 4).

Including only pupils associated with the reopening scenarios reduces the system to between 21% and 100% of all schools and 35% and 66% of all households (Table 1).



**Figure 3** A graph of the school to school contact network of all 21,608 schools. Nodes show schools connected by edges with widths that indicate the number of unique contact pairs between the schools. Red nodes show secondary schools, blue nodes show primary schools.



**Figure 4** Degree distributions of the UK school contact network (All schools). Purple shows the weighted degree - the total number of contact pairs with adjacent schools. Green shows the unweighted degree - total number of adjacent schools.

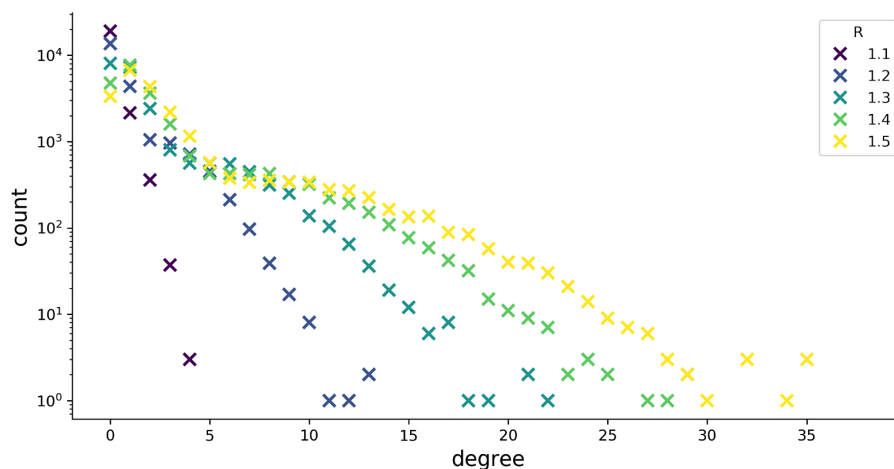
scenario	schools	% of all Schools	households	% of all households
1	17953	83	1728173	37
2	21438	99	2211384	45
3	19982	93	1926090	39
4	21480	100	2381729	48
5	17984	83	3267414	66
6	4555	21	2627640	53
<b>all schools</b>	<b>21583</b>	<b>100</b>	<b>4927163</b>	<b>100</b>

**Table 1** The number of schools and households represented in each school reopening scenario.

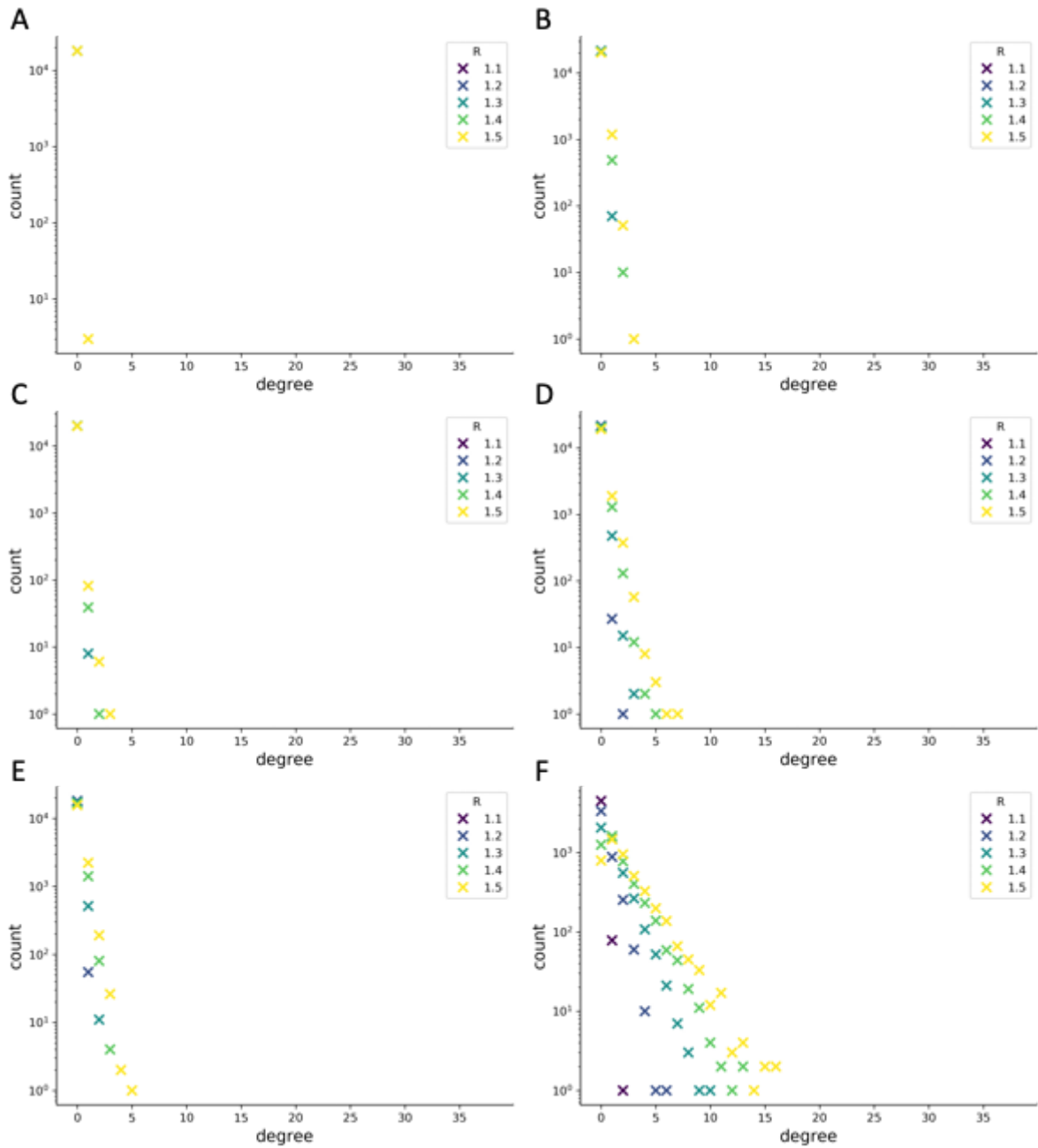
### *Degree distributions of the transmission probability network*

With all schools fully open, the mean weighted degree of the transmission probability network varied between 0.42 for R of 1.1, to 3.6 for an R of 1.5. The school with the highest degree varied between 4.7 to 35.5 for R of 1.1 and 1.5 respectively.

When the network was modified to only include pupils from certain years the mean degrees reduced (Figure 6). Scenario 1 (Reception and years 1 and 6) had the lowest mean degree for all values of R, whilst scenario 6 (opening secondary schools only) had the highest, 0.26 - 2.6 across values of R 1.1 to 1.5. After Scenario 6, Scenario 5 (primary schools only) had the highest mean degree, between 0.05 and 0.45. Scenarios 2 - 4, which all combined some partial opening of primary and secondary schools, had relatively similar degree distributions to that of opening only, but all years of, primary schools (Table 2). Of these, scenario 3 (R, 1, 6 and 12) had the lowest mean degree for each value of R, between 0.01 and 0.15.



**Figure 5** Weighted degree distribution (expected number of schools infected by each school) of the transmission probability network with all schools fully open for R values of 1.1 to 1.5.



**Figure 6** Weighted degree distribution (expected number of schools infected by each school) of the transmission probability network for each of the reopening scenarios considered for R values of 1.1 to 1.5. A - F show scenarios 1 - 6 respectively.

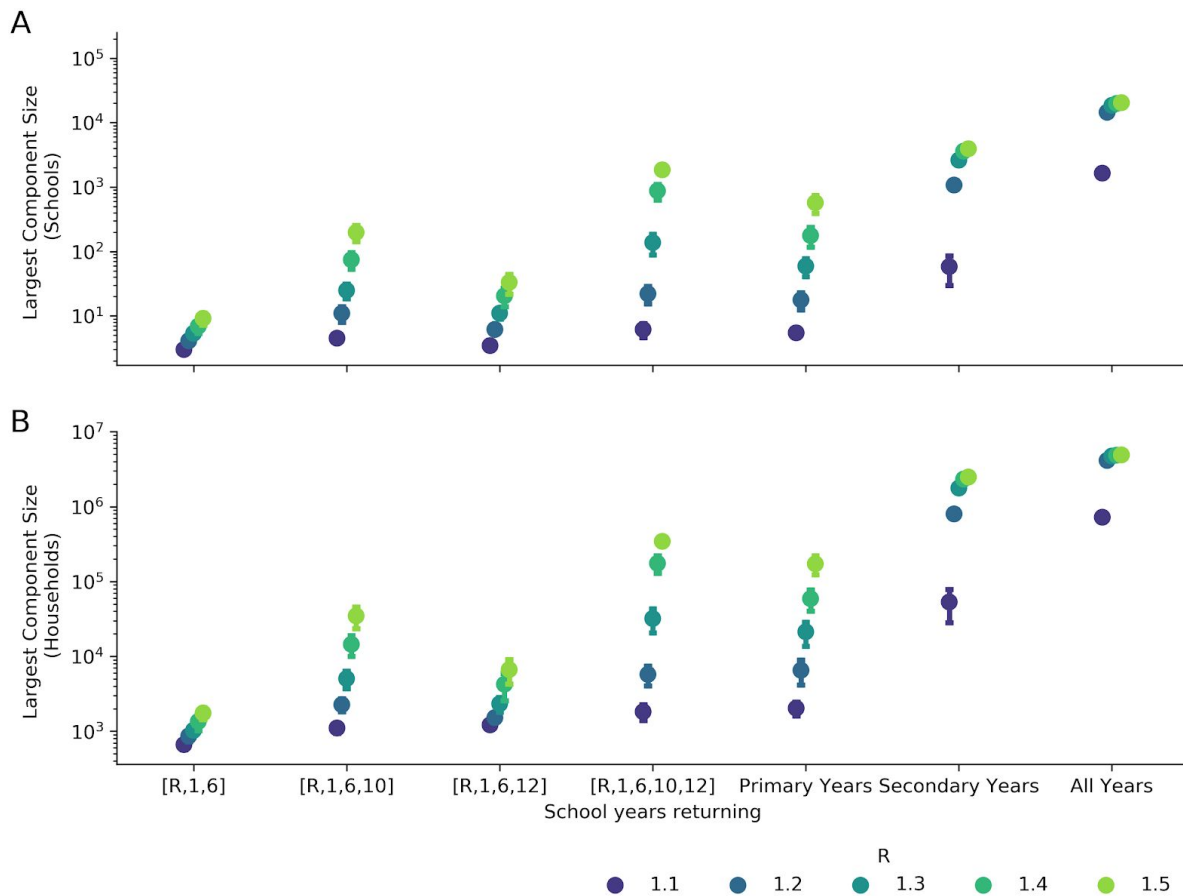


Scenario	R	weighted degree		
		median	mean	max
1	1.1	0.00	0.01	0.13
	1.2	0.02	0.02	0.39
	1.3	0.03	0.05	0.67
	1.4	0.04	0.07	0.95
	1.5	0.06	0.09	1.20
2	1.1	0.02	0.03	0.27
	1.2	0.06	0.10	0.87
	1.3	0.12	0.18	1.60
	1.4	0.17	0.26	2.37
	1.5	0.23	0.34	3.11
3	1.1	0.01	0.01	0.30
	1.2	0.02	0.04	0.98
	1.3	0.04	0.08	1.82
	1.4	0.07	0.11	2.71
	1.5	0.09	0.15	3.58
4	1.1	0.02	0.04	0.66
	1.2	0.08	0.12	2.13
	1.3	0.14	0.22	3.91
	1.4	0.21	0.33	5.74
	1.5	0.27	0.44	7.50
5	1.1	0.02	0.05	0.60
	1.2	0.08	0.15	1.46
	1.3	0.14	0.26	2.61
	1.4	0.22	0.36	3.87
	1.5	0.29	0.45	5.10
6	1.1	0.19	0.26	2.21
	1.2	0.60	0.81	6.46
	1.3	1.08	1.43	10.60
	1.4	1.57	2.04	13.98
	1.5	2.02	2.60	16.91

**Table 2** Median, mean and maximum weighted degree on the transmission probability network for each scenario for R values between 1.1 and 1.5.

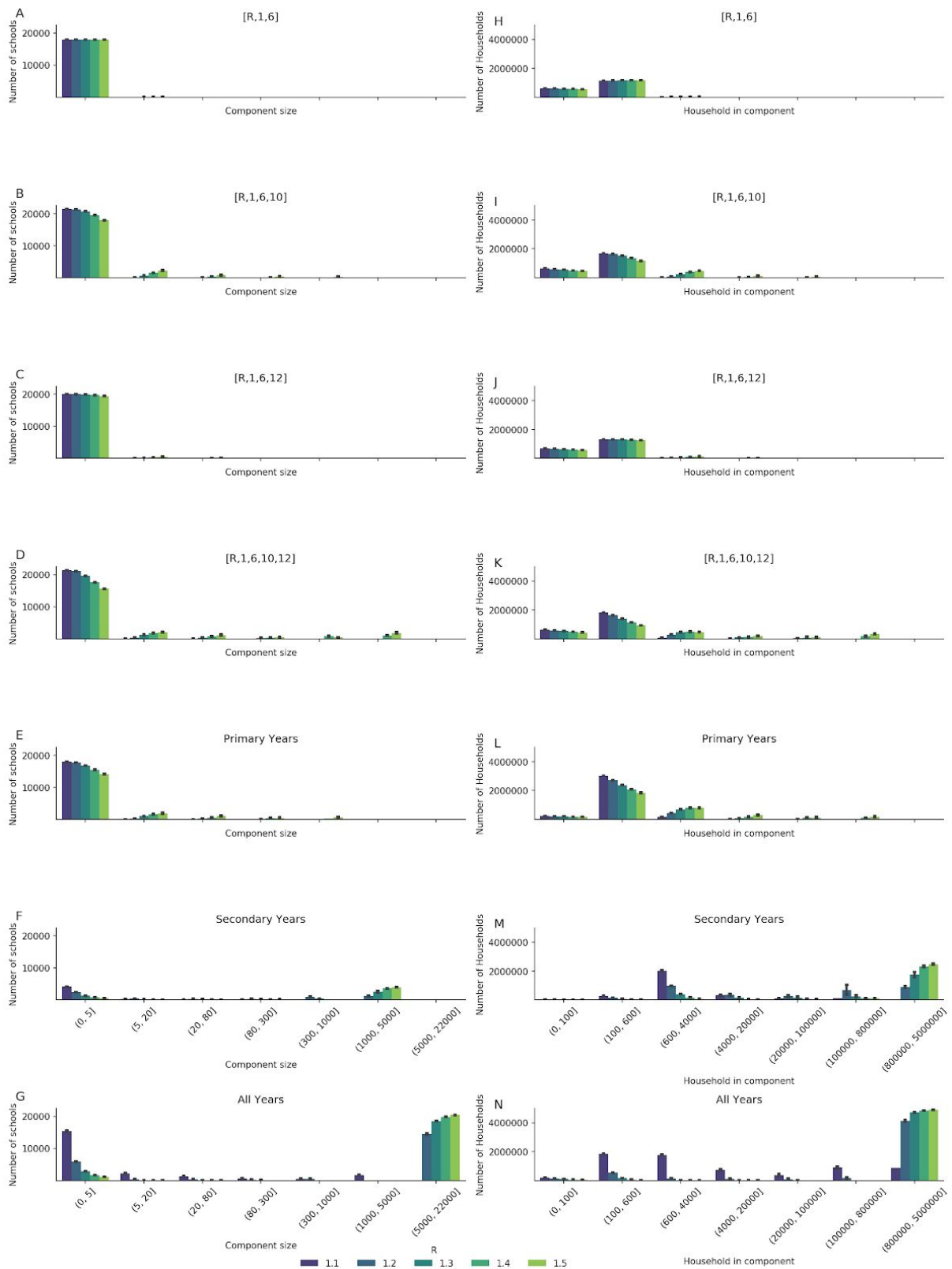
**Connected components of binary transmission networks**

The largest connected component of the realisations of the binary transmission networks, that is, the number of schools in the largest connected part of the network, increased with R for each scenario (figure 7). For scenario 1 (Reception, Year 1, Year 6) the largest components simulated were not much greater than 10 schools, and there were very few exceeding 10 schools in each realisation (figure 8), these connected components typically represented less than 1000 households in total.



**Figure 7** Largest component of the binary transmission networks generated from transmission probability networks for school reopening scenarios. A) the number of schools in each largest connected component, B) the number of households with children attending a school in each largest connected component.

Adding secondary school years 10 or 12 to the network (Scenario 4) increased the largest connected component size considerably, although comparable to scenario 1 at R of 1.1, the largest connected components for realisations at R of 1.5 reached several hundred schools and tens of thousands of households. The distribution of component sizes was also shifted up a number of components larger than 100 schools, however the majority of schools remain in components of less than 10 schools.



**Figure 8** Connected component distribution of the binary transmission networks generated for school reopening scenarios and R values of 1.1 to 1.5 (indicated by colour). By school (A-G), i.e. the number of schools in a component size in each bin and by household (H-N)

As well as having a greater largest connected component, there were more large connected components for scenario 2 (year 10) than scenario 3 (year 12). Including both (scenario 4) increased the largest component again, reaching an average of around 2000 schools for an R of 1.5 associated with over 200,000 households, the majority of schools and households remained in small components of less than 10 schools, containing fewer than 1000 households. i.e. the number of schools in a component with the number of households associated within each bin.

Opening all years of primary schools (scenario 5) was comparable to, but did not create as large components as scenario 4. Scenario 6, reopening secondary schools, vastly exceeded all other scenarios where a large proportion (>25%) of secondary schools were included in the largest connected component for R of 1.3 or greater, where the number of schools in the largest component of more than 3000 schools, associated with over 2 million households, nearly 50% of the households of children to schools represented in the network.

## **Discussion**

We constructed a network of schools connected by households using individual level pupil data provided by DfE. By analysing the network we have shown the relative potential for schools to play a role in local and longer range transmission under various scenarios of reopening. We analysed the network to assess the potential for additional risk of transmission through the community caused by reopening schools.

The degree distribution of the transmission probability network and analysis of connected components of realisations of binary transmission networks show that scenarios with a small selection of school years may be possible with only a small risk of transmission between schools. In particular the analysis highlights the difference in risk posed by secondary schools relative to primary schools, where reopening even a small subset of secondary school years (year 10 and 12) increases the connectivity between schools considerably, whereas opening all primary schools resulted in lower connectivity in the network. Furthermore opening only secondary schools results in the highest connectivity of all the scenarios evaluated. If it is assumed that secondary school pupils will be better able to practice physical distancing than younger children in primary schools, this could act to counterbalance this risk.

Although the value of R has a substantial impact on the size of the largest connected component, there was little impact on the results for the vast majority of schools' component sizes. There was however some impact on the weighted degree distribution suggesting that the virus may spread more quickly across the connected components with higher within-school R.

This analysis is subject to a number of important limitations: The network can only capture transmission in schools and households between school age children and aims to provide insight into the capacity for transmission within schools and households only. The data from which the network is constructed included only state funded schools in England with children coded as school years Reception to Year 13 in official data, addition of independent schools would increase the size and possibly the connectivity of the network, however only 7% of children in England attend an independent school so the impact would be expected to be

marginal. The quantification of the probability of transmission between schools assumed that each school outbreak reached its natural final size - this should be unlikely before interventions, such as targeted school or class closure are introduced. Transmission within schools was assumed to be homogeneous and transmission rate was assumed to be the same in all schools. This approach was chosen to maintain the parsimony of the approach, modelling internal transmission dynamics of individual schools would considerably increase the scope of the framework. Transmission between children was assumed to be common with a reproduction number in schools greater than 1. There is currently no clear scientific consensus as to how susceptible and infectious school age children are regarding SARS-COV-2<sup>10</sup>. Our estimates are conservative in the sense that there if children were less susceptible or infectious than adults, R within schools could be less than 1 - in the event that evidence arises to the contrary our work should be considered in light of this.

The analysis provides insight into the potential for school based and household based contacts between children to combine to create long chains of transmission which could result in infections within many thousands of households. We highlight that the impact of reintroduction of children into school on transmission varies substantially between the tested scenarios. Introducing primary school children had much lower risk of transmission between schools than secondary schools children. We also highlight that maintaining restrictions on contact between children within schools to ensure a low within school reproduction number may be highly influential, as the rate of transmission between schools increases rapidly with internal R. Further analysis using this network may provide more precise guidance, particularly on reactive school closure strategies in the event of detecting a school outbreak.

---

<sup>10</sup> Russell M Viner, Oliver T Mytton, Chris Bonell, G.J. Melendez-Torres, Joseph L Ward, Lee Hudson, Claire Waddington, James Thomas, Simon Russell, Fiona van der Klis, Jasmina Panovska-Griffiths, Nicholas G Davies, Robert Booy, Rosalind Eggo; Susceptibility to and transmission of COVID-19 amongst children and adolescents compared with adults: a systematic review and meta-analysis medRxiv 2020.05.20.20108126; doi: <https://doi.org/10.1101/2020.05.20.20108126>