

# Visualizing the network structure of Covid-19 in Singapore

Tod Van Gunten  
tvangun@ed.ac.uk

## Abstract

Many infectious diseases like Covid-19/SARS-CoV-2 spread through pre-existing social networks. While network models address the implications of micro-level interaction patterns for disease transmission, epidemiologists and social scientists know relatively little about the meso structure of virus spread. Meso structure refers to the pattern of disease spread at a higher level of aggregation, that is, between infection clusters corresponding to organizations, locales, and events. This paper visualizes this meso structure using publicly available contact tracing data from Singapore. Visualization shows that one highly central infection cluster appears to have generated on the order of 7-8 infection chains, amounting to 60% of non-imported cases during the period considered. However, no other cluster generated more than two infection chains. This heterogeneity suggests that network meso-structure is highly consequential for epidemic dynamics.

Many infectious diseases spread through pre-existing social networks (???). Similarly, Covid-19/SARS-CoV-2 propagates in families, workplaces, social gatherings and other localized settings. Epidemiological models incorporate estimates of the degree distribution and homophily of social contacts in order to evaluate the impact of networks on disease spread (?). Similarly, research suggests that a small number of infected individuals result in a majority of subsequent SARS-CoV-2 infections, generating infection clusters known as super-spreading events (?). These findings begin to illustrate the impact of network micro-structure, or the local patterns of social interaction that affect viral spread. However, we still know little about what we might call the network meso-structure of viral disease transmission.

Network meso-structure refers here to connections between individuals and clusters that affect transmission between bounded social settings. (In contrast, macro-structure encompasses the spread of disease between cities and nations). Hypothetically, just as individuals may transmit the virus to a heterogeneous number of subsequent individuals, a cluster could give rise to a varied number of infection chains and subsequent clusters. A super-spreading event in a social setting connected to many other settings could potentially produce many more infections than an event occurring in a less well-connected setting. Infection meso-structure likely depends on social, cultural and policy aspects of the social settings generating clusters, including socio-spatial segregation, mobility, employment patterns, public health measures and many other factors.

I visualize the network meso-structure of SARS-CoV-2 infection in Singapore using publicly available contact tracing data. Singapore is an interesting and perhaps unique case; public health authorities were unusually successful in controlling the virus through stringent measures, but a substantial outbreak emerged in dormitories housing the city's guest worker population in March 2020. Disease transmission was thus heavily affected by the particular inequalities of Singaporean society (?). An important question is therefore how the virus spread to many different socially bounded settings (dormitories) with limited impact on the non-migrant population.

Figure 1 represents the six largest network components (see supplement for data details). Square nodes represent infected individuals while circles represent clusters of multiple infections. For visual clarity, individuals whose only tie is to one cluster are omitted as nodes and represented by scaling

cluster diameter to the number of associated individuals. Ties represent personal contact between individuals or associations of individuals with clusters. Collectively these five components account for 77% of non-imported Covid-19 cases detected between March 25 and April 19, 2020.

Panel A shows the largest connected component, which accounts for 60% of non-imported cases, while panel B shows additional components with  $\geq 20$  nodes. The largest component consists of a highly central cluster (cluster 1) surrounded by 7–8 contagion chains. Cluster 1 is a major construction site cited by health authorities as an important hub of transmission. Based on the dates of cluster origination, it appears that all but one of these chains post-date cluster 1 (the remaining chain is ambiguous); thus this cluster appears to be a critical pathway of virus transmission. Notably, while the overwhelming majority of Covid-19 patients were guest workers, this cluster was primarily non-guest worker.

An interesting feature of this structure is that the branching of infection chains appears limited, even during a period of rapid spread. Only cluster 1 generated on the order of 7–8 infection chains; most other clusters form part of 1–2 infection chains. Cluster 1 appears to be a kind of ‘super-super spreading’ event that not only produced many infections, but produced many infection chains. Other clusters are ‘only’ super-spreading events producing many infections, but few infection chains. The two largest clusters (clusters 2 and 3), both dormitories, did not produce multiple subsequent infection chains. This may reflect the effects of public health interventions. An implication is that patterns of connection between social settings (e.g. between a construction site and dormitories) are consequential for viral diffusion. Public health interventions to remove interactions between cluster 1 and other sites might have been effective in preventing an outbreak. In summary, this visualization shows suggests that the spread of a virus such as SARS-Cov-2 is structured by meso-level interaction patterns in addition to micro-social dynamics.

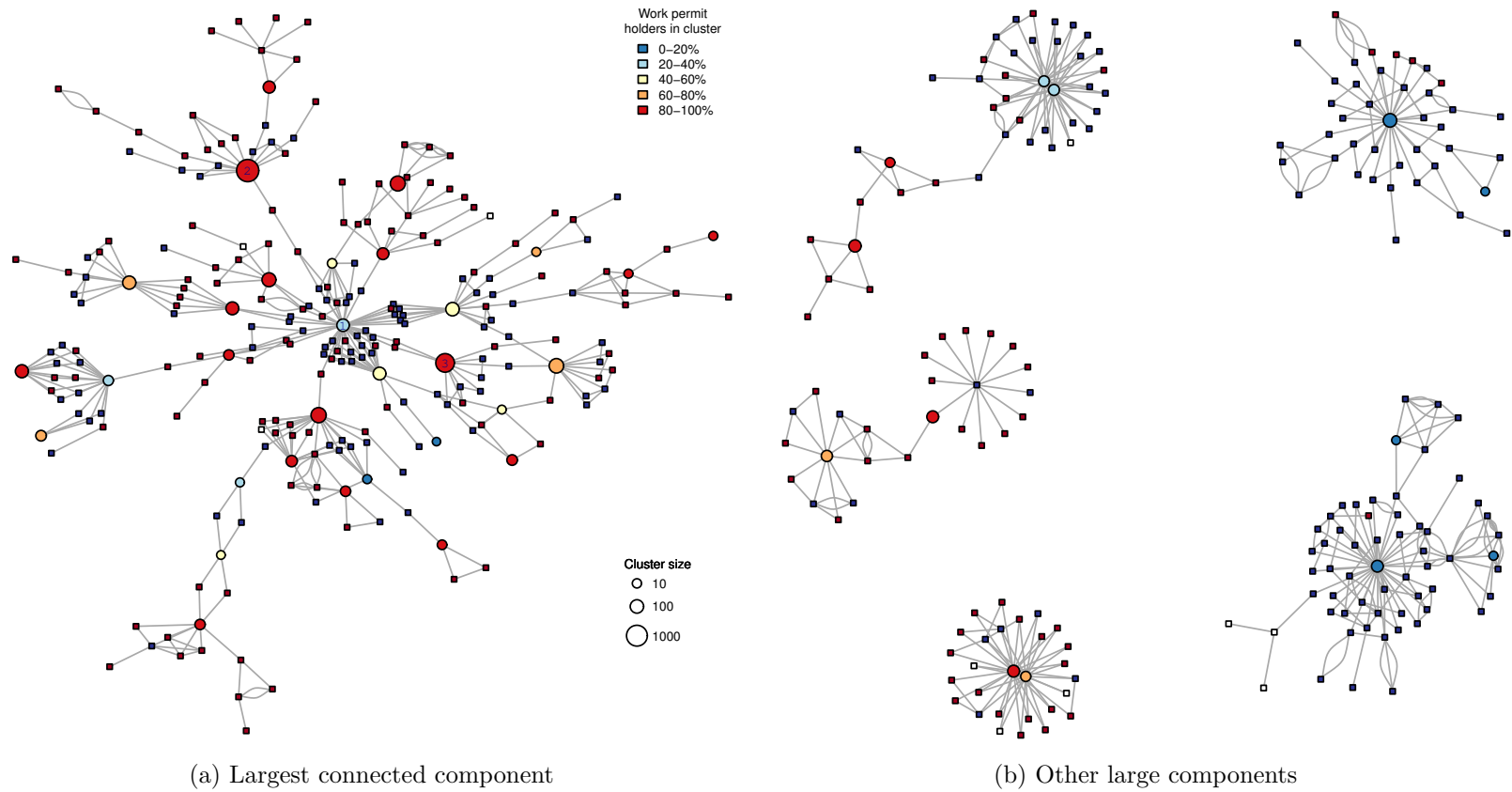


Figure 1: SARS-CoV2-infection network in Singapore: connections between individuals and clusters  
 Red squares represent Work Permit holders; blue squares represent other individuals. Circles represent clusters scaled to the percentage of work permit holders by color and cluster size by radius.

# 1 The Singaporean context

Singapore is an interesting case in which to examine SARS-CoV-2 transmission networks because of interaction patterns in the population and seemingly unique pandemic dynamics. The first case of Covid-19 in Singapore was identified on January 23, 2020. Local spread of SARS-CoV-2 was slow for two months, but a substantial outbreak began in late March. While viral spread during February and early March occurred in Singaporean society at large (termed ‘community transmission’ by local public health authorities), the later outbreak between April and September 2020 occurred primarily among work permit holders (guest workers) residing in dormitory settings. Bangladesh and India are the two most common national origins of guest workers. These settings are highly densely occupied, with as many as twenty workers occupying a single room, creating a context where viral transmission was difficult to control. As of September 15, 54,248 Covid-19 cases were identified in dormitory settings, amounting to 94% of all cases in Singapore.<sup>1</sup> Numerous international media reports drew attention to the inequalities of the guest worker system and the plight of workers confined to cramped settings (??). Remarkably, as of December 1, 2020 Singapore had seen only 29 Covid-19 deaths.<sup>2</sup>

# 2 Data sources and network construction

From the first detected case of Covid-19 infection through April 19, 2020 the Singapore Ministry of Health (MoH) provided daily press releases with anonymized case-level information, including links between individuals and infection clusters (press releases are available from [moh.gov.sg/covid-19](http://moh.gov.sg/covid-19)). These data are derived from the MoH contact tracing program. Information on links and cluster associations was frequently updated after initial announcements. I include all cases, link and cluster updates through April 19, 2020. Subsequent to April 19, public individual-level information is more limited and does not permit identification of links between cases and clusters. Individual-level variables include test date, gender, nationality, migration status, and travel history. In addition to individual cases, MoH officials identified 79 clusters prior to April 19: clusters are multiple infections associated with a location or event. About one third of clusters are dormitories.

**Supplementary information for ‘Visualizing the network structure of Covid-19 in Singapore’**

MoH press releases and accompanying appendices provide two sources of network data: links between individuals (i.e. known social contacts) and associations of individuals with clusters. A link between individuals means that this pair are identified as contacts through the MoH contact tracing program (though we cannot know with certainty that infection occurred through this contact). A link between an individual and a cluster means that MoH officials determined that an individual was physically present at a location or event determined to constitute a cluster. Visualizations adopt the Kamada-Kawai layout.

For visual clarity, figure 1 omits all individuals whose only known tie is to a cluster. These individuals are represented by the scaling of cluster diameter to (log of) cluster size. In other words, infected individuals are represented as nodes only if they have ties to two individuals, two clusters, or one individual and one cluster. The reason for this choice is illustrated in supplementary figure S1, which shows the main component represented in figure 1b retaining all individuals as nodes. Visual clarity is reduced and these nodes are largely redundant to the scaling of cluster size. (Figure S1 scales cluster diameter to the square root of cluster size rather than the logarithm for reasons for visual clarity).

### 3 Network meso-structure

As noted in the main document, previous research examines network properties in order to understand the dynamics of disease spread. For example, sociological research has examined the structure of romantic networks in order to draw inferences about the spread of sexually transmitted diseases (?). But most epidemiological research focuses on local network properties such as the degree distribution. Epidemiologists stress that the reproduction number  $r_t$  (the number of infections resulting from a source) likely varies widely, with the tail of the distribution reflecting ‘super-spreading’ events: individuals who transmit the virus to many others (?). Less is known about the number of infection chains resulting from such events and the length of chains between clusters. I conceptualize this meso-structure as the pattern of infection relations between clusters, including both clusters themselves and the infection chains between them. This meso structure is not a feature of virus biology, but rather of the social interactions, mobility patterns and physical

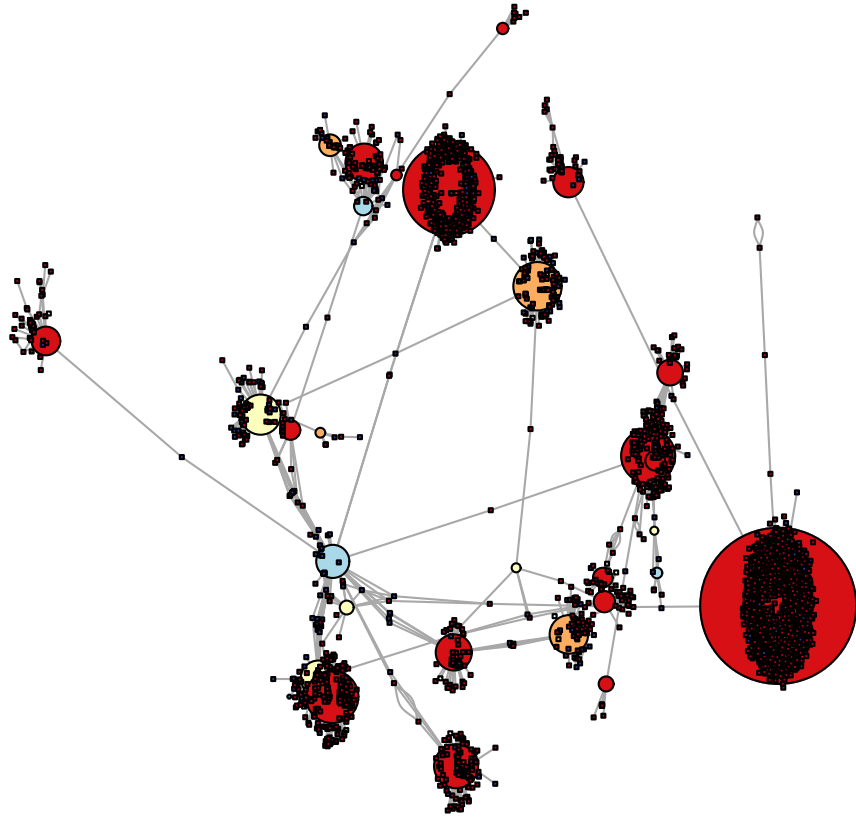


Figure S1: Main component of transmission network, without removal of individuals connected to only one cluster

Supplementary information for 'Visualizing the network structure of Covid-19 in Singapore'

proximity that create the potential for virus transmission. Meso structure is also distinct from macro structure, which involves the still higher level of aggregation of transmission between cities, regions and countries [e.g.] (?).

One possible model of transmission meso-structure is a multiple hub-and-spoke model, according to which each infection cluster generates a varying number of infection chains (i.e. the cluster-level degree distribution), each of which generates additional clusters which in turn seed further chains. This model of spread could arise if local network properties (e.g. heterogeneous degree distribution and clustering) affect transmission locally, but spread between clusters is not substantially constrained by meso-level social interaction patterns. Notably figure 1 does not conform to this model: rather, while cluster 1 generated 7–8 infection chains, other clusters only generated 1–2 infection chains. This observation is supported by an inter-cluster reduction of the network. This network consists of all clusters in the main component; ties between clusters are infection chains through individuals of any length (supplement figure S2). This meso-structure may reflect interaction patterns between these apparently later clusters, for example, the relative infrequency of interactions between individuals in different dormitory settings. It is beyond the scope of this paper to account for this network structure; rather the goal is to draw attention to the social interaction patterns at higher levels of aggregation that may impact infectious disease spread.

## 4 Broader network context

It is important to emphasize that the network represented in figure 1 consists of the largest connected components within the transmission network, *not* the universe of all connected individuals, or all individuals with a positive Covid–19 test. MoH assigned one of three statuses to each case: imported (i.e. an individual appeared to have contracted Covid–19 abroad), linked (an individual had social contact with another infected individual, or was associated with a cluster) or unlinked (no known connections to other cases and clusters. Appendix table 1 provides a summary of this broader network context.

Overall, a very high-proportion of cases are linked to 6 network components (including the largest component) — 92% of cases initially classified as ‘local linked’ by MoH and 75% of all local



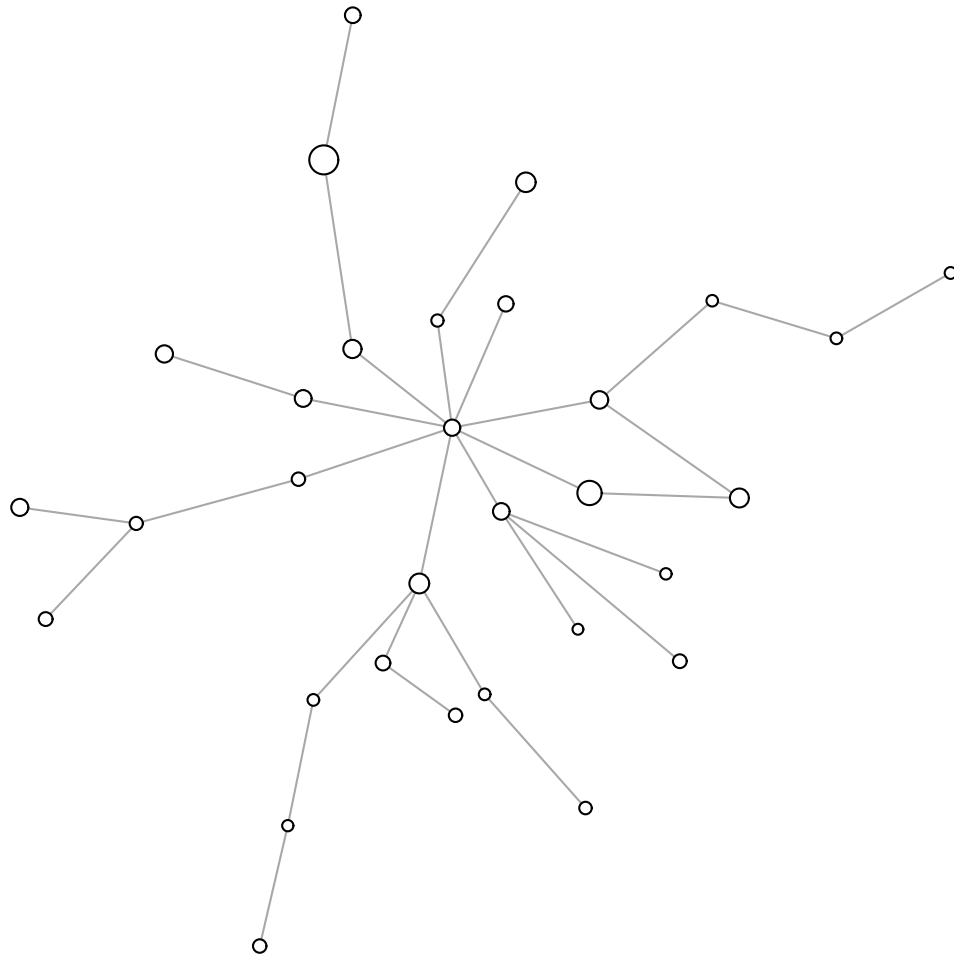


Figure S2: Inter-cluster reduction of main component (see main text for explanation). Position of clusters corresponds to position in main text figure 1, panel A.

Table S1: Distribution of network connectivity

Type	Main component		Component n $\geq$ 10		Whole network	
	N	%	N	%	N	%
Jan 23, 2020 - Apr. 19, 2020						
Imported	1	0%	17	3%	173	31%
Local linked	3010	73%	3801	92%	4118	100%
Local unlinked	457	24%	739	39%	898	47%
All local	3466	57%	4539	75%	5015	83%
Mar 25, 2020 - April 19, 2020						
Imported	1	1%	10	6%	48	27%
Local linked	3010	76%	3679	93%	3932	100%
Local unlinked	456	25%	726	40%	854	47%
All local	3465	60%	4404	77%	4785	83%

cases. (The discrepancy is due to cases initially classified ‘unlinked’ by MoH but for which links were later identified). This has two main implications. First, it suggests that MoH contact tracing was highly effective at detecting contacts between cases and clusters. Second, the fact that such a high percentage of linked cases are connected in just six components suggests that these data are informative about transmission meso-structure. Furthermore, 76% of cases occurring between the first case associated with the main component (detected March 25) and the end of the data on April 19 are associated with the main component. This suggests that the main component is a reliable representation of the meso-structure of Covid-19 transmission during this period. Figure S2 shows the network including all connected nodes to illustrate this broad network context.

Nevertheless, due to unobserved transmission events, this network is necessarily an approximation. Approximately 6% of cases with known connections are located in small disconnected components (e.g. dyads), and 20% of non-imported cases have no known connections to other cases or clusters. Assuming that SARS-CoV-2 transmission always occurs between specific individuals (even if in fleeting interactions), these cases represent missing ties (unobserved transmission events) in the network. However, inclusion of these unobserved events in the network is unlikely to have a substantial impact on the observed meso-structure.

In order for these unobserved transmission events to substantially alter the network meso-structure, unobserved ties would have to exhibit a very high level of network autocorrelation:

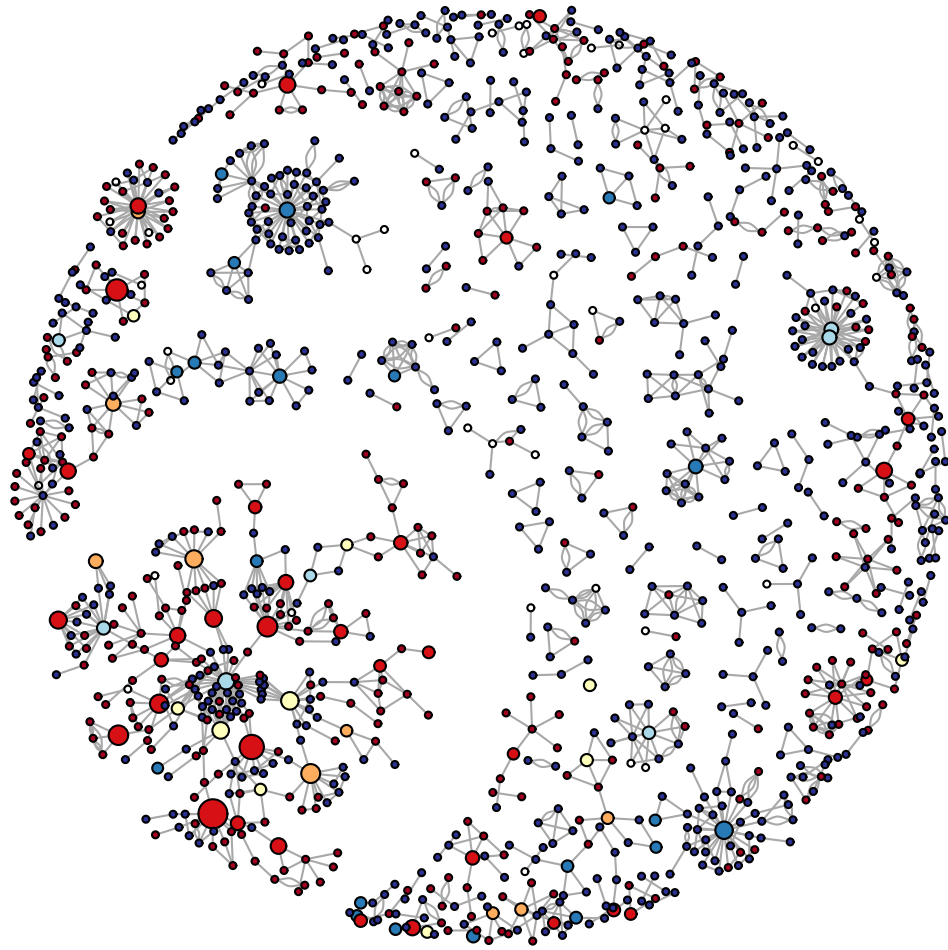


Figure S3: Representation of all connected individuals and clusters

multiple disconnected nodes would need to be extremely likely to have connections to other nodes in order for these to amount to a structurally meaningful effect. This level of autocorrelation is unlikely to exist in practice.

For example, according to media reports, MoH officials believed that a cluster in panel B — localized in a mall — was the source for cluster 1 (panel A), with transmission moving from mall employees to workers (?). While this transmission chain is not observed in the network data (likely due to undetected transmission), the structure of transmission within the main component suggests that this is likely a short path through a small number of individuals. This would not alter the overall picture of Covid-19 meso-structure in panel A.

## References

- Adam, Dillon C., Peng Wu, Jessica Y. Wong, Eric H.Y. Lau, Tim K. Tsang, Simon Cauchemez, Gabriel M. Leung, and Benjamin J. Cowling. 2020. “Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong.” *Nature Medicine* 26:1714–1719.
- Bearman, Peter S., James Moody, and Katherine Stovel. 2004. “Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks.” *American Journal of Sociology* 110:44–91.
- Chinazzi, Matteo, Jessica T. Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun, Cécile Viboud, Xinyue Xiong, Hongjie Yu, M. Elizabeth Halloran, Ira M. Longini, and Alessandro Vespignani. 2020. “The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak.” *Science* 368:395–400.
- Lai, Linette. 2020. “Coronavirus: Mustafa Centre believed to be starting point for hundreds of cases.”
- Mossong, Joël, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W. John Edmunds. 2008. “Social contacts and mixing patterns relevant to the spread of infectious diseases.” *PLoS Medicine* 5:0381–0391.
- Pellis, Lorenzo, Frank Ball, Shweta Bansal, Ken Eames, Thomas House, Valerie Isham, and Pieter Trapman. 2015. “Eight challenges for network epidemic models.” *Epidemics* 10:58–62.
- Salathé, Marcel, Maria Kazandjieva, Jung Woo Lee, Philip Levis, Marcus W. Feldman, and James H. Jones. 2010. “A high-resolution human contact network for infectious disease

transmission.” *Proceedings of the National Academy of Sciences of the United States of America* 107:22020–22025.

Stack, Megan K. 2020. “A Sudden Coronavirus Surge Brought Out Singapore’s Dark Side.” *New York Times Magazine* .

Tan, Yvette. 2020. “Covid-19 Singapore: A ‘pandemic of inequality’ exposed.” *BBC News* .

Teo, You Yenn. 2018. *This is What Inequality Looks Like*. Singapore: Ethos Books.