

1 **Crowding and the epidemic intensity of COVID-19 transmission**

2

3 Benjamin Rader^{1,2}, Anjalika Nande³, Ben Adlam³, Alison L. Hill³, Robert C. Reiner⁴ David M. Pigott⁴,
4 Bernardo Gutierrez^{5,6}, open COVID-19 data working group[#], John S. Brownstein^{1,7}, Marcia C. Castro⁸,
5 Huaiyu Tian⁹, Oliver G. Pybus^{5,10,\$}, Samuel V. Scarpino^{11,\$}, Moritz UG Kraemer^{1,5,7,\$}

6

- 7 1. Computational Epidemiology Lab, Boston Children’s Hospital, Boston, United States
8 2. Department of Epidemiology, Boston University School of Public Health, Boston, United States
9 3. Program for Evolutionary Dynamics, Harvard University, Cambridge, United States
10 4. Department of Health Metrics, University of Washington, Seattle, United States
11 5. Department of Zoology, University of Oxford, Oxford, United Kingdom
12 6. School of Biological and Environmental Sciences, Universidad San Francisco de Quito USFQ,
13 Quito, Ecuador
14 7. Harvard Medical School, Boston, United States
15 8. Department of Global Health and Population, Harvard T.H. Chan School of Public Health,
16 Boston, United States
17 9. State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System
18 Science, Beijing Normal University, Beijing, China
19 10. Department of Pathobiology and Population Science, The Royal Veterinary College, London,
20 United Kingdom
21 11. Network Science Institute, Northeastern University, Boston, United States

22

23 ^{\$}correspondence should be addressed to moritz.kraemer@zoo.ox.ac.uk and s.scarpino@northeastern.edu
24 and oliver.pybus@zoo.ox.ac.uk

25 [#]Members of the working group are listed at the end of the manuscript

26

27 **Abstract**

28 **The COVID-19 pandemic is straining public health systems worldwide and major non-**
29 **pharmaceutical interventions have been implemented to slow its spread¹⁻⁴. During the initial phase**
30 **of the outbreak the spread was primarily determined by human mobility^{5,6}. Yet empirical evidence**
31 **on the effect of key geographic factors on local epidemic spread is lacking⁷. We analyse highly-**
32 **resolved spatial variables for cities in China together with case count data in order to investigate**
33 **the role of climate, urbanization, and variation in interventions across China. Here we show that**
34 **the epidemic intensity of COVID-19 is strongly shaped by crowding, such that epidemics in dense**
35 **cities are more spread out through time, and denser cities have larger total incidence. Observed**
36 **differences in epidemic intensity are well captured by a metapopulation model of COVID-19 that**
37 **explicitly accounts for spatial hierarchies. Densely-populated cities worldwide may experience more**
38 **prolonged epidemics. Whilst stringent interventions can shorten the time length of these local**
39 **epidemics, although these may be difficult to implement in many affected settings.**

40

41 Predicting the epidemiology of the COVID-19 pandemic is a central priority for guiding epidemic
42 responses around the world. China has undergone its first epidemic wave and, remarkably, cities across
43 the country are now reporting few or no locally-acquired cases⁸. Analyses have indicated that that the
44 spread of COVID-19 from Hubei to the rest of China was driven primarily by human mobility⁶ and the
45 stringent measures to restrict human movement and public gatherings within and among cities in China
46 have been associated with bringing local epidemics under control⁵. Key uncertainties remain as to which
47 geographic factors drive local transmission dynamics and affect the intensity of transmission of COVID-
48 19. For respiratory pathogens, “epidemic intensity” (*i.e.*, the peakedness of the number of cases through
49 time, or the shortest period during which the majority of cases are observed) varies with increased indoor
50 crowding, and socio-economic and climatic factors⁹⁻¹³. Epidemic intensity is minimized when incidence
51 is spread evenly across weeks and increases as incidence becomes more focused in particular days
52 (**Figure 1C**, see a detailed description of how epidemic intensity is defined in Ref. ⁹). In any given
53 location, higher epidemic intensity requires a larger surge capacity in the public health system¹⁴,
54 especially for an emerging respiratory pathogen such as COVID-19¹⁵.

55

56 China provides richly detailed epidemiological time series^{2,16,17} across a wide range of geographic
57 contexts, hence the epidemic there provides an opportunity to evaluate the role of factors in shaping the
58 intensity of local epidemics. We use detailed line-list data from Chinese cities^{18,19}, climate and population
59 data, local human mobility data from Baidu, and timelines of outbreaks responses and interventions to

60 identify drivers of local transmission in Chinese cities, with a focus on epidemic intensity among
61 provinces in China.

62

63 To explore the impact of urbanization, temperature, and humidity, we used daily incidence data of
64 confirmed COVID-19 cases (date of onset) aggregated at the prefectural level ($n = 293$) in China.

65 Prefectures are administrative units that typically have one urban center (**Figure 1**). We aggregate

66 individual level data that were collected from official government reports¹⁷. Epidemiological data in each

67 prefecture were truncated to exclude dates before the first and after the last day of reported cases. The

68 shape of epidemic curves varied between prefectures with some showing rapid rises and declines in cases

69 and others showing more prolonged epidemics (**Figure 1A**). We estimate epidemic intensity for each

70 prefecture from these data by calculating the inverse Shannon entropy of the distribution of incident

71 cases⁹. We define the incidence distribution p_{ij} for a given city to be the proportion of COVID-19 cases

72 during epidemic wave j that occurred on day i . The inverse Shannon entropy of incidence for a given

73 prefecture and year is then given by $v_j = (-\sum_i p_{ij} \log p_{ij})^{-1}$. Because v_j is a function of the disease

74 incidence curve in each location, rather than of absolute incidence values, it is invariant under differences

75 in overall reporting rates among cities or overall attack rates. Population counts for each prefecture were

76 extracted from a 1 km x 1 km gridded surface of the world utilizing administrative-2 level cartographic

77 boundaries.

78

79 Within each prefecture, we calculate Lloyd's index of mean crowding^{9,20} treating the population count of

80 each pixel as an individual unit (**Methods, Figure 1B and C**). The term 'mean crowding' used here is a

81 specific metric that summarizes both, population density and how density is distributed across a

82 prefecture (patchiness). Values on the resulting index above the mean pixel population count within each

83 prefecture suggest a spatially-aggregated population structure (**Methods**). For example, Guangzhou has

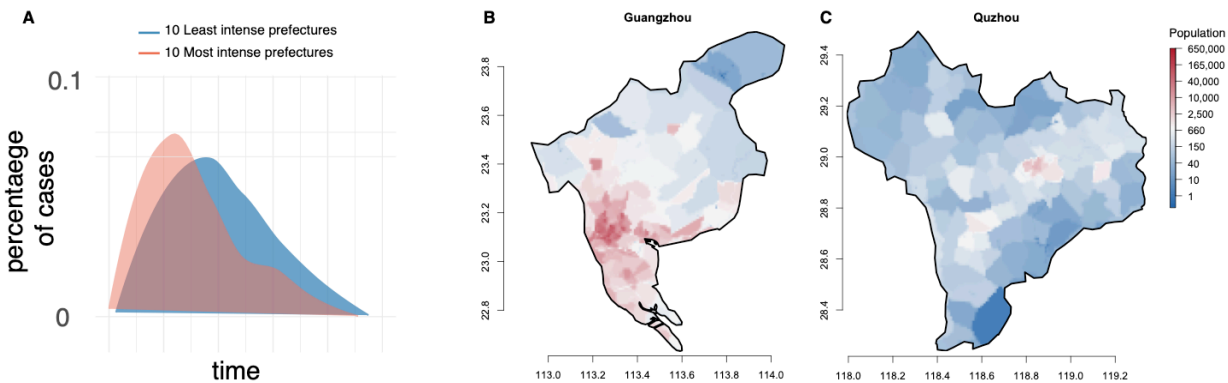
84 high values of crowding whilst Quzhou which has a more evenly distributed population in its prefecture

85 (**Figure 1B and C**). Using the centroid of each prefecture we calculate daily mean temperature and

86 specific humidity; these values are subsequently averaged over each prefecture's reporting period

87 (**Methods**). We performed log-linear regression modeling to determine the association between epidemic

88 intensity with the socio-economic and environmental variables (**Methods**).



89

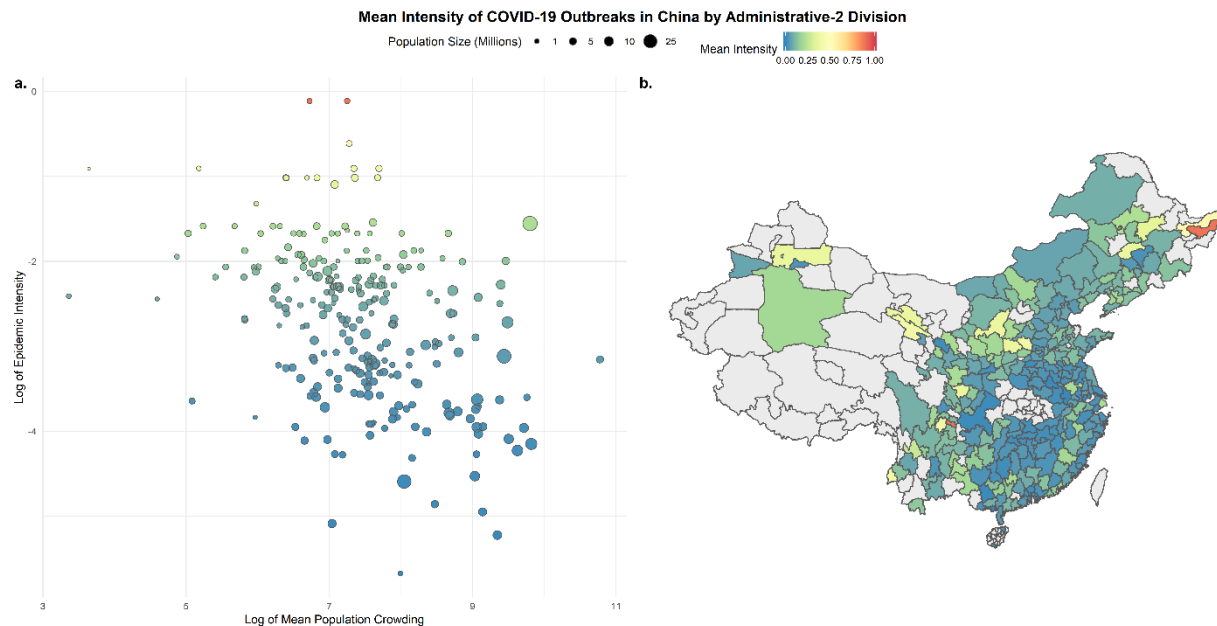
90

91 **Figure 1: Maps of crowding in prefectures in China.** A) shows epidemic curves that are normalized to
92 show the percentage of cases that are occurring at each given day. The 10 most intense (red) prefectures
93 are shown versus the 10 least intense (blue). B) An example of a prefecture with high levels of crowding
94 (Guangzhou, Guangdong Province), versus (C) a prefecture where populations are more equally
95 distributed across the prefecture (Quzhou, Zhejiang Province). The colour scale illustrates the number of
96 inhabitants per grid cell (1km x 1km).

97

98 We found that epidemic intensity is significantly negatively correlated with mean population crowding
99 and varies widely across the country (**Figure 2**, Extended Data Table 1, p-value < 0.001). Our
100 observation contrasts those expected from simple and classical epidemiological models where it would be
101 expected to see more intensity in crowded areas^{21,22}. We hypothesize that the mechanism that underlies
102 the more crowded cities experience less intense outbreaks because crowding enables more widespread
103 and sustained transmission between households leading incidence to be more widely distributed in time
104 (see section below for detailed simulation, **Methods**). Population size, mean temperature, and mean
105 specific humidity were all significant but their correlation coefficients were much smaller (**Extended**
106 **Data Table 1**). A multivariate-model was able to explain a large fraction of the variation in epidemic
107 intensity across Chinese cities ($R^2 = 0.54$). We perform sensitivity analysis to account for potential noise
108 in the city level incidence distribution (Extended Data Fig. 1).

109



110
111 **Figure 2: Crowding and the intensity of transmission of COVID-19 in China. a)** negative association
112 *between log of epidemic intensity, as measured by inverse Shannon entropy (Methods), and log*
113 *population crowding, as measure by Lloyd's mean crowding (Methods). Lower intensity and therefore*
114 *prolonged epidemics in larger cities. The size of the points indicate the size of the population in each city,*
115 **b)** *Map of epidemic intensity in China at the prefecture level. Darker colours indicate lower intensity and*
116 *lighter colours higher intensity. Grey prefectures had not enough reported cases, no cases or were not*
117 *included in analyses (Hubei Province).*

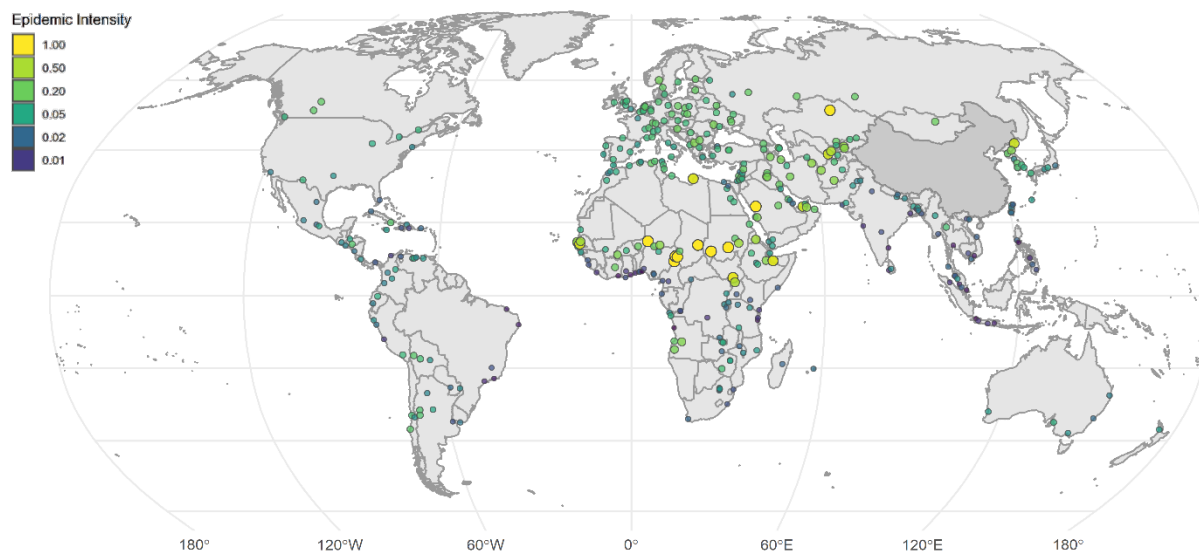
118
119 One key uncertainty in previous applications of models of epidemic intensity was the contribution of
120 disease importation(s) on the shape of the epidemic⁹. Due to the unprecedented scale of human mobility
121 restrictions imposed in China, the fact that the early epidemic was effectively from a single source,
122 coupled with the availability of real-time data on mobility, we can evaluate the impact of these
123 restrictions on the epidemic intensity relative to the local dynamics. To do so, we performed a univariate
124 analysis (**Extended Data Table 1**) and found that human mobility explained 14% of the variation in
125 epidemic intensity. This further supports earlier findings that COVID-19 had already spread throughout
126 much of China prior to the cordon sanitaire of Hubei province and that the pattern of seeding potentially
127 modulates epidemic intensity^{6,23}. These findings are in agreement with previous work on other pathogens
128 (measles, influenza) which showed that once local epidemics are established case importation becomes
129 less important in determining epidemic intensity²⁴.

130

131 To evaluate the potential impact of variability of intensity on the peak incidence and total incidence we
132 performed a simple linear regression. We found that peak incidence was correlated with epidemic
133 intensity (locations that had high intensity also had more cases at the peak). Total incidence, however,
134 was larger in areas with lower estimated intensity, which is intuitive as crowded areas have longer
135 epidemics that affect more people (**Extended Data Table 2**). This suggests that measures taken to
136 mitigate the epidemic may need to be enforced more strictly in smaller cities to lower the peak incidence
137 (flatten the curve) but conversely may not need to be implemented as long. Furthermore, with lower total
138 incidence in small cities, the risk of resurgence may be elevated due to lower population immunity. There
139 is urgent need to collect serological evidence to provide a full picture of attack rates across the world²⁵.

140
141 Using our model trained on cities in China we extrapolated epidemic intensity to cities across the world
142 (**Figure 3**). Figure 3 shows the distribution of epidemic intensity in 380 urban centers. Cities in yellow
143 are predicted to have higher epidemic intensity relative to those in blue (a full list is provided in
144 **Extended Data Table 3**). Small inland cities in sub-Saharan Africa had high predicted epidemic intensity
145 and may be particularly prone to experience large surge capacity in the public health system²⁶. In general,
146 coastal cities had lower predicted intensity and larger and more prolonged predicted epidemics. Global
147 predictions of epidemic intensity in cities rely on fitted relationships of the first epidemic curve from
148 Chinese prefectures and therefore need to be interpreted with extreme caution.

149



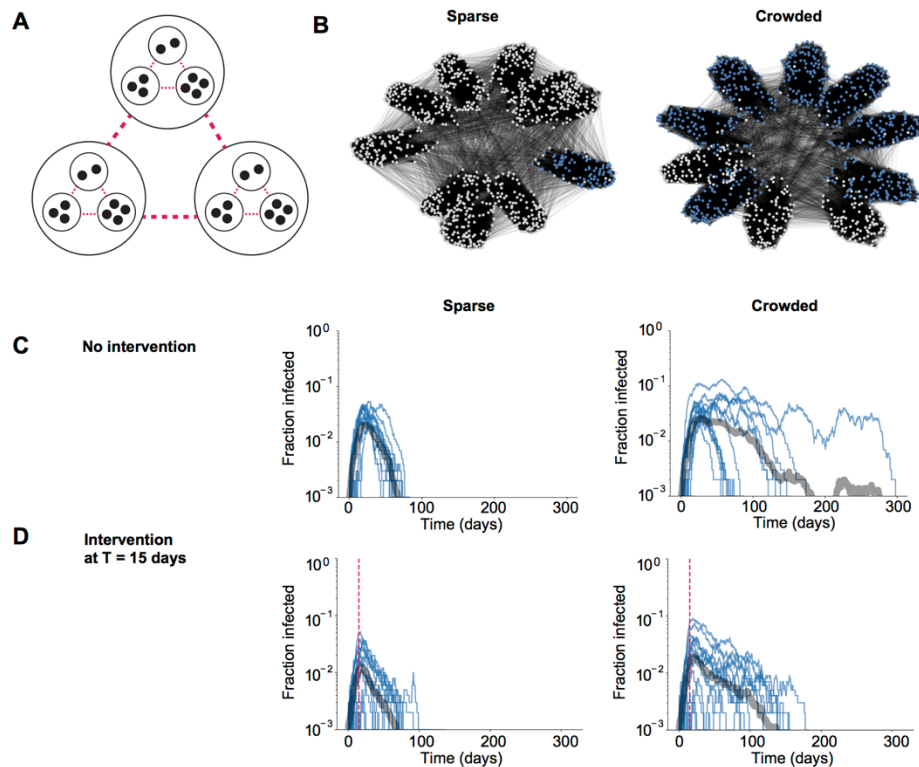
150
151 **Figure 3: Predicted epidemic intensities vary across 380 global cities. Darker colours represent low**
152 **epidemic intensity and lighter colours represent high epidemic intensity. Estimates were generated using**
153 **the full model (Model 5) fitted to epidemic curves in Chinese cities (Extended Data Table 1). A full list of**

154 *epidemic intensities can be found in the Extended Data Table 3. Epidemic intensity is a measure of*
155 *peakedness of epidemics and does not reflect the expected number of cases (Methods).*

156

157 To understand the mechanism responsible for our finding that outbreaks in crowded cities were lower
158 intensity (*i.e.* more spread out in time), we simulated stochastic epidemic dynamics in different types of
159 populations. Simple, well-mixed transmission models where contact rates are higher in crowded regions
160 were not consistent with our findings, since they predict crowded regions would have more intense and
161 higher-peaked outbreaks. To capture more realistic contact patterns, we created hierarchically-structured
162 populations²⁷ where individuals had high rates of contact within their households (households are defined
163 broadly and could represent care homes, hospitals, prisons, etc.), lower rates with individuals from other
164 households but within the same “neighborhoods”, and relatively rare contact with other individuals in the
165 same prefecture (**Figure 4A**). Assumptions are consistent with reports that the majority of onward
166 transmission occurred in households^{2,28}. We assumed spread between prefectures was negligible once an
167 outbreak started. In this scenario, “sparse” prefectures often had more intense, short-term outbreaks that
168 were isolated to certain neighborhoods, while “crowded” prefectures could have drawn-out, low intensity
169 outbreaks that jumped between the more highly-connected “neighborhoods” (**Figures 4B and C**). These
170 outbreaks had larger final size than those in less-crowded areas (**Figure 4C**) which likely is related to
171 large overdispersion in the reproduction number of COVID-19^{29,30} where local outbreaks can reach their
172 full potential due to the availability of contacts. We also considered outbreak dynamics in sparse and
173 crowded prefectures under strong social distancing measures, which is likely to be the scenario occurring
174 across China during most of the time captured by our study and certainly after January 23, 2020². If social
175 distancing reduces non-household contacts by the same relative amount in all prefectures, there will be
176 more contacts remaining in crowded areas, since baseline contact rates are higher. In this case, it may take
177 much longer for the infection to die out post-intervention in crowded areas (**Figure 4D**), leading to a
178 lower intensity outbreak with larger final size, as seen in our data (**Figure 1C**).

179



180

181 **Figure 4: Mechanisms generating less intense epidemics in crowded populations.** A) Schematic of a
182 hierarchically-structured population model consisting of households and “neighborhoods” within a
183 prefecture. Transmission is more likely among contacts connected at lower spatial levels. Crowded
184 populations have stronger connections outside the household, and interventions reduce the strength of
185 these connections in both populations (pink lines). B) - C) Simulated outbreak dynamics in the absence of
186 interventions in crowded vs sparse populations. For the networks in (B), blue nodes are individuals who
187 were eventually infected by the end of the outbreak. In (C), individual realizations are shown with thin
188 blue lines and the average in the thick grey line. D) Simulated outbreak dynamics in the presence of
189 strong social distancing measures in crowded vs. sparse populations. The intervention was implemented
190 at day 15 (pink line) and led to a 75% reduction in contacts.

191

192 Spatial covariates and particularly crowding are important parameters to consider in the assessment of
193 epidemics across the world. Crowded cities tend to be more prolonged due to increased crowding and the
194 higher potential for transmission chains to persist (*i.e.*, in denser environments there is higher potential for
195 two randomly selected hosts in a population to attain spatiotemporal proximity sufficient for COVID-19
196 transmission). Our findings confirm previous work on epidemic intensity of transmission of influenza in
197 cities⁹ albeit the mechanism for influenza is likely driven by the accumulation of immunity rather than the
198 specific network structure of individuals. More generally, our work provides empirical support for the

199 role of spatial organization in determining infectious disease dynamics and the limited capacity of *cordon*
200 *sanitaires* to control local epidemics^{27,31}. We were unable to test more specific hypotheses about which
201 interventions may have impacted the intensity of transmission within and between cities. Further, even
202 though humidity was negatively associated with epidemic intensity it did not explain the majority of the
203 variation and more work will be needed to find causal evidence for the effect of humidity on transmission
204 dynamics of COVID-19. Therefore, maps showing epidemic intensity in cities outside China (**Figure 3**)
205 should be interpreted with particular caution.

206

207 Currently, non-pharmaceutical interventions are the primary control strategy for COVID-19. As a result,
208 public health measures are often focused on ‘flattening the curve’ to lower the risk of essential services
209 running out of capacity. We show that spatial context, especially crowding, can result in a higher risk of
210 intensive epidemics in less crowded, comparatively rural areas. Therefore, it will be critical to view non-
211 pharmaceutical interventions through the perspective of crowding (*i.e.*, how does an intervention reduce
212 the circle of contacts of an average individual) in cities across the world. Specifically, cities in sub-
213 Saharan Africa have high predicted epidemic intensities that will likely overwhelm already stressed health
214 care systems.

215

216 **References**

- 217 1. Fraher, E. P. *et al.* Ensuring and Sustaining a Pandemic Workforce. *N. Engl. J. Med.*
218 *NEJMp2006376* (2020). doi:10.1056/NEJMp2006376
- 219 2. Leung, K., Wu, J. T., Liu, D. & Leung, G. M. First-wave COVID-19 transmissibility and severity
220 in China outside Hubei after control measures, and second-wave scenario planning: a modelling
221 impact assessment. *Lancet* **6736**, (2020).
- 222 3. Ji, Y., Ma, Z., Peppelenbosch, M. P. & Pan, Q. Potential association between COVID-19 mortality
223 and health-care resource availability. *Lancet Glob. Heal.* **8**, e480 (2020).
- 224 4. Rosenbaum, L. Facing Covid-19 in Italy — Ethics, Logistics, and Therapeutics on the Epidemic’s
225 Front Line. *N. Engl. J. Med.* *NEJMp2005492* (2020). doi:10.1056/NEJMp2005492
- 226 5. Tian, H. *et al.* An investigation of transmission control measures during the first 50 days of the
227 COVID-19 epidemic in China. *Science* eabb6105 (2020). doi:10.1126/science.abb6105
- 228 6. Kraemer, M. U. G. *et al.* The effect of human mobility and control measures on the COVID-19
229 epidemic in China. *Science* **21**, eabb4218 (2020).
- 230 7. Lipsitch, M., Swerdlow, D. L. & Finelli, L. Defining the Epidemiology of Covid-19 — Studies
231 Needed. *N. Engl. J. Med.* **382**, 1194–1196 (2020).
- 232 8. World Health Organization (WHO). Coronavirus disease 2019 (COVID-19) Situation Report - 71.

- 233 (2020).
- 234 9. Dalziel, B. D. *et al.* Urbanization and humidity shape the intensity of influenza epidemics in U.S.
235 cities. *Science* **362**, 75–79 (2018).
- 236 10. Shaman, J., Pitzer, V. E., Viboud, C., Grenfell, B. T. & Lipsitch, M. Absolute humidity and the
237 seasonal onset of influenza in the continental United States. *PLoS Biol.* **8**, (2010).
- 238 11. Gog, J. R. *et al.* Spatial transmission of 2009 pandemic influenza in the US. *PLoS Comput. Biol.*
239 **10**, e1003635 (2014).
- 240 12. Shaman, J. & Kohn, M. Absolute humidity modulates influenza survival, transmission, and
241 seasonality. *Proc. Natl. Acad. Sci.* **106**, 3243–3248 (2009).
- 242 13. Chetty, R. *et al.* The association between income and life expectancy in the United States, 2001-
243 2014. *JAMA - J. Am. Med. Assoc.* **315**, 1750–1766 (2016).
- 244 14. Crawford, J. M. *et al.* Laboratory Surge Response to Pandemic (H1N1) 2009 Outbreak, New York
245 City Metropolitan Area, USA. *Emerg. Infect. Dis.* **16**, 8–13 (2010).
- 246 15. Grasselli, G., Pesenti, A. & Cecconi, M. Critical Care Utilization for the COVID-19 Outbreak in
247 Lombardy, Italy. *JAMA* **19**, NEJMoa2002032 (2020).
- 248 16. Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected
249 Pneumonia. *N. Engl. J. Med.* NEJMoa2001316 (2020). doi:10.1056/NEJMoa2001316
- 250 17. Xu, B. *et al.* Epidemiological data from the COVID-19 outbreak, real-time case information. *Sci.*
251 *Data* **7**, (2020).
- 252 18. Xu, B. *et al.* Epidemiological data from the COVID-19 outbreak, real-time case information.
253 *figshare* (2020). doi:10.6084/m9.figshare.11949279
- 254 19. Xu, B. & Kraemer, M. U. G. Open access epidemiological data from the COVID-19. *Lancet*
255 *Infect. Dis.* **3099**, 30119 (2020).
- 256 20. Lloyd, M. 'Mean Crowding'. *J. Anim. Ecol.* **36**, 1 (1967).
- 257 21. May, R. M. & Anderson, R. M. Spatial heterogeneity and the design of immunization programs.
258 *Math. Biosci.* **72**, 83–111 (1984).
- 259 22. Anderson, R. M. & May, R. M. *Infectious diseases of humans: dynamics and control*. (Oxford
260 University Press, 1991).
- 261 23. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel
262 coronavirus (COVID-19). *Science* **3221**, 2020.02.14.20023127 (2020).
- 263 24. Bjørnstad, O. N. & Grenfell, B. T. Hazards, spatial transmission and timing of outbreaks in
264 epidemic metapopulations. *Environ. Ecol. Stat.* **15**, 265–277 (2008).
- 265 25. Lipsitch, M., Swerdlow, D. L. & Finelli, L. Defining the Epidemiology of Covid-19 — Studies
266 Needed. *N. Engl. J. Med.* NEJMp2002125 (2020). doi:10.1056/NEJMp2002125

- 267 26. Martinez-Alvarez, M. *et al.* COVID-19 pandemic in west Africa. *Lancet Glob. Heal.* **2019**, 2019–
268 2020 (2020).
- 269 27. Watts, D. J., Muhamad, R., Medina, D. C. & Dodds, P. S. Multiscale, resurgent epidemics in a
270 hierarchical metapopulation model. *Proc. Natl. Acad. Sci.* **102**, 11157–11162 (2005).
- 271 28. Aylward, Bruce (WHO); Liang, W. (PRC). Report of the WHO-China Joint Mission on
272 Coronavirus Disease 2019 (COVID-19). **2019**, 16–24 (2020).
- 273 29. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of
274 individual variation on disease emergence. *Nature* **438**, 355–359 (2005).
- 275 30. Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: a mathematical
276 modelling study. *Lancet Infect. Dis.* **3099**, 1–7 (2020).
- 277 31. Sattenspiel, L. Simulating the Effect of Quarantine on the Spread of the 1918–19 Flu in Central
278 Canada. *Bull. Math. Biol.* **65**, 1–26 (2003).
- 279 32. Ramshaw, R. E. *et al.* A database of geopositioned Middle East Respiratory Syndrome
280 Coronavirus occurrences. *Sci. Data* **6**, 318 (2019).
- 281 33. Doxsey-Whitfield, E. *et al.* Taking Advantage of the Improved Availability of Census Data: A
282 First Look at the Gridded Population of the World, Version 4. *Pap. Appl. Geogr.* **1**, 226–234
283 (2015).
- 284 34. Reiczigel, J., Lang, Z., Rózsa, L. & Tóthmérész, B. Properties of crowding indices and statistical
285 tools to analyse parasite crowding data. *J. Parasitol.* **91**, 245–252 (2005).
- 286 35. Wade, M. J., Fitzpatrick, C. L. & Lively, C. M. 50-year anniversary of Lloyd’s “mean crowding”:
287 Ideas on patchy distributions. *J. Anim. Ecol.* **87**, 1221–1226 (2018).
- 288 36. Florczyk, A. *et al.* GHS-UCDB R2019A - GHS Urban Centre Database 2015, multitemporal and
289 multidimensional attributes. *Eur. Comm. Jt. Res. Cent.* (2019).
- 290
- 291

292 **Methods**

293 Epidemiological data

294 No officially reported line list was available for cases in China. We use a standardised protocol³² to
295 extract individual level data from December 1st, 2019 - March 30th, 2020. Sources are mainly official
296 reports from provincial, municipal or national health governments. Data included basic demographics
297 (age, sex), travel histories and key dates (dates of onset of symptoms, hospitalization, and confirmation).
298 Data were entered by a team of data curators on a rolling basis and technical validation and geo-
299 positioning protocols were applied continuously to ensure validity. A detailed description of the
300 methodology is available¹⁷. Lastly, total numbers were matched with officially reported data from China
301 and other government reports.

302

303 Estimating epidemic intensity

304 Epidemic intensity was estimated for each prefecture by calculating the inverse Shannon entropy of the
305 distribution of COVID-19 cases. Shannon entropy was used to fit time series of other respiratory
306 infections (influenza)⁹. The Shannon entropy of incidence for a given prefecture and year is then given by
307 $v_j = (-\sum_i p_{ij} \log p_{ij})^{-1}$. Because v_j is a function of incidence distribution in each location rather than
308 raw incidence it is invariant under differences in overall reporting rates between cities or attack rates. We
309 then assessed how mean intensity $v \propto \sum_j v_j$ varied across geographic areas in China.

310

311 Proxies for COVID-19 interventions

312 Real-time measures of human mobility were extracted from the Baidu Qianxi web platform to estimate
313 the proportion of daily movement between the city of Wuhan to Hubei and 30 other Chinese provinces.
314 Relative mobility volume was available from January 2, 2020 to January 25, 2020 and averaged across
315 these dates to create a single measure of relative flows from Wuhan. This data was only available at the
316 province level, so each individual prefecture inherited the relative mobility of its higher-level province.
317 Baidu's mapping service is estimated to have a 30% market share in China and more data can be found^{5,6}.

318

319 Data on drivers of transmission of COVID-19

320 Prefecture-specific population counts and densities were derived from the 2020 Gridded Population of
321 The World, a modeled continuous surface of population estimated from national census data and the
322 United Nations World Population Prospectus³³. Population counts are defined at a 30 arc-second
323 resolution (approximately 1 km x 1 km at the equator) and extracted within administrative-2 level
324 cartographic boundaries defined by the National Bureau of Statistics of China. Lloyd's mean crowding,

325 $\frac{[\sum_i(q_i-1)q_i]}{\sum_i q_i}$, was estimated for each prefecture where q_i represents the population count of each non-zero
326 pixel within a prefecture's boundary and the resulting value estimates an individual's mean number of
327 expected neighbors^{9,34}.

328
329 Daily temperature (°F), relative humidity (%) and atmospheric pressure (Pa) at the centroid of each
330 prefecture was provided by The Dark Sky Company via the Dark Sky API and aggregated across a
331 variety of data sources. Specific humidity (kg/kg) was then calculated using the R package, `humidity`¹².
332 Meteorological variables for each prefecture were then averaged across the entirety of the study period.

333

334 Statistical analysis

335 We normalized the values of epidemic intensity between 0 and 1, and for all non-zero values fit a
336 Generalized Linear Model (GLM) of the form:

337

$$338 \log(Y_j) \sim \beta_0 + \beta_1 \log(C_j) + \beta_2 q_j + \beta_3 \log(P_j) + \beta_3 f_j + \beta_4 R_j$$

339

340 where for each prefecture j , Y is the scaled Shannon-diversity measure of epidemic intensity derived from
341 the COVID-19 time series, C is Lloyd's Index of Mean Crowding^{20,35}, q is the mean specific humidity
342 over the reporting period in kg/kg, P is the estimated population count and f is the relative population
343 flows from Wuhan to each prefecture's higher level province. To account for the length of the epidemic
344 period in each city we calculate R as the number of reporting days.

345

346 Projecting epidemic intensity in cities around the world

347 We selected 380 urban centers from the European Commission Global Human Settlement Urban Centre
348 Database and their included cartographic boundaries³⁶. To ensure global coverage, up to the five most
349 populous cities in each country were selected from the 1,000 most populous urban centers recorded in the
350 database. Population count, crowding, and meteorological variables were then estimated following
351 identical procedures used to calculate these variables in the Chinese prefectures. Weather measurements
352 were averaged over the 2-month period starting on February 1, 2020.

353

354 The parameters from the model of epidemic intensity predicted by humidity, crowding and population
355 size (see Table 1, Model 6) were used to estimate relative intensity in the 380 urban centers. Predicted
356 values of epidemic intensity that fell outside the original covariate space $[0,1]$ ($n=7$) were set to 1. A full
357 list of predicted epidemic intensities can be found in the Supplementary Information.

358

359 Sensitivity analyses

360 The inverse Shannon entropy metric may be sensitive to noise in incidence distribution. For example, the
361 noisier the incidence distribution the higher the epidemic intensity. To the extent that noise is elevated in
362 small populations (due to demographic stochasticity for instance) intensity also tends to be higher in
363 smaller populations, even if they have the same underlying shape to their epidemic curve. Lloyd's mean
364 crowding also varies strongly with population size. Therefore, some of the observed relationship between
365 intensity and crowding may be due to (possibly independent) statistical scaling of both intensity and
366 crowding with population size. We therefore perform sensitivity analysis to test if cities that are more
367 crowded than expected for their size have lower epidemic intensities than expected for their size. We
368 calculate 'excess intensity' as the residuals on a regression of $\log(\text{epidemic intensity}) \sim \log(\text{pop})$; 'excess
369 crowding' as the residuals on a regression of $\log(\text{crowding}) \sim \log(\text{pop})$ and plot the relationship between
370 excess intensity and excess crowding' (Extended Data Figure 1).

371

372 Simulating epidemic dynamics

373 We simulated a simple stochastic SIR model of infection spread on weighted networks created to
374 represent hierarchically-structured populations. Individuals were first assigned to households using the
375 distribution of household sizes in China (data from UN Population Division, mean 3.4 individuals).
376 Households were then assigned to "neighborhoods" of ~ 100 individuals, and all neighborhood members
377 were connected with a lower weight. A randomly-chosen 10% of individuals were given "external"
378 connections to individuals outside the neighborhood. The total population size was $N=1000$. Simulations
379 were run for 300 days and averages were taken over 20 iterations. The SIR model used a per-contact
380 transmission rate of $\beta=0.15/\text{day}$ and recovery rate $\gamma=0.1/\text{day}$. For the simulations without interventions,
381 the weights were $w_{HH} = 1$, $w_{NH} = 0.01$, and $w_{EX} = 0.001$ for the "crowded" prefecture and $w_{EX} = 0.0001$ for
382 the "sparse" prefecture. For the simulations with interventions, the household and neighborhood weights
383 were the same but we used $w_{EX} = 0.01$ for the "crowded" prefecture and $w_{EX} = 0.001$ for the "sparse"
384 prefecture. The intervention reduced the weight of all connections outside the household by 75%.

385

386

387 **Acknowledgements:** The authors thank Kathryn Cordiano for her statistical assistance. BR
388 acknowledges funding from Google.org. MUGK acknowledges funding from European Commission
389 H2020 program (MOOD project) and a Branco Weiss Fellowship. OGP and HT acknowledge funding
390 from the Oxford Martin School. ALH and AN acknowledge funding from the US National Institutes of
391 Health (DP5OD019851). The funding bodies had no role in study design, data collection and analysis,
392 preparation of the manuscript, or the decision to publish. All authors have seen and approved the
393 manuscript.

394

395 **Author contributions** MUGK, SVS, OGP conceived the research. BR, ALH, AN, BA, SVS, MUGK
396 analysed the data. MUGK wrote the first draft of the manuscript. All authors contributed to interpretation
397 of results and manuscript writing.

398

399 **Competing interests:** The authors declare no competing interests.

400

401 **Data availability:** We collated epidemiological data from publicly available data sources (news articles,
402 press releases and published reports from public health agencies) which are described in full here¹⁷. All
403 the epidemiological information that we used is documented in the main text, the extended data, and
404 supplementary tables.

405

406 **Code availability:** The code is available from this link: [tbc](#) and the simulation code is available from
407 here: <https://github.com/alsnhll/SIRNestedNetwork>

408