

Real-time Dissection and Forecast of Infection Dynamics during a Pandemic

Steven Schulz^{a,*}, Richard Pastor^a, Cenk Koyuncuoglu^a, Forrest W. Crawford^{b,c,d,e}, Detlef Zernick^a, André Karch^f, and Sten Rüdiger^a

^a*Machine Learning and Health Unit, Department of Engineering, NET CHECK GmbH, Berlin, Germany*

^b*Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA*

^c*Department of Statistics and Data Science, Yale University, New Haven, CT, USA*

^d*Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA*

^e*Yale School of Management, New Haven, CT, USA*

^f*Institute of Epidemiology and Social Medicine, Westfälische Wilhelms-Universität Münster, Münster, Germany*

Abstract

Pandemic preparedness requires institutions, including public health authorities and governments, to detect, survey and control outbreaks. To maintain an accurate, quantitative and up-to-date picture of an epidemic crisis is key. For SARS-CoV-2, this was mostly achieved by ascertaining incidence numbers and the effective reproductive number (R_{eff}), which counts how many people an infected person is likely to infect on average. These numbers give strong hints on past infection dynamics in a population but fail to clearly characterize current and future dynamics as well as potential effects of pharmaceutical and non-pharmaceutical interventions. We show that, by using and combining infection surveillance and population-scale contact statistics, we can obtain a better understanding of the drivers of epidemic waves and the effectiveness of interventions. This approach can provide a real-time picture, thus saving not only many lives by quickly allowing adaptation of the health policies but also alleviating economic and other burdens if an intervention proves ineffective. We factorize R_{eff} into contacts and relative transmissibility: Both signals can be used, individually and combined, to identify driving forces of an epidemic, monitoring and assessing interventions, as well as projecting an epidemic's future trajectory. Using data for SARS-CoV-2 and Influenza from 2019 onward in Germany, we provide evidence for the usefulness of our approach. In particular, we find that the effects from physical distancing and lockdowns as well as vaccination campaigns are dominant.

Keywords: epidemiology, contact networks, transmissibility, real-time measurement & forecast

1. Introduction

1 Infectious diseases represent serious threats to
2 an ever increasingly connected humankind, on par
3 with e.g. natural disasters and infrastructure fail-
4 ures. Epidemic preparedness – the ability to pre-
5 dict and mitigate future epidemic outbreaks – has
6 thus risen to one of the most pressing challenges in

8 modern societies and recently focused a wealth of
9 research efforts building on a variety of data [1] in
10 response to awareness elicited by the SARS-CoV-2
11 pandemic [2].

12 Epidemic dynamics are shaped at the crossroads
13 of human and viral driving forces: a pathogen's re-
14 productive cycle, defining its relative transmission
15 rate upon physical proximity between individuals
16 with full or partial susceptibility, as well as hu-

*Corresponding author: steven.schulz@netcheck.de.

17 man behaviour, via the frequency of transmission-
18 prone contacts between individuals itself [3]. Criti-
19 cal events such as the emergence of fitter mutants or
20 collective shifts in human activity patterns set the
21 pace for new epidemic waves. Real-time monitoring
22 of these forces during an epidemic, whether it is fu-
23 eled mostly by increased contact levels or changes
24 in relative transmissibility, is of paramount value
25 for epidemic forecasting as well as the ability to set
26 up informed, targeted mitigation strategies and es-
27 timating the effects of (non-)pharmaceutical health
28 policies [4].

29 Using SARS-CoV-2 and Influenza as key exam-
30 ples of airborne transmissible contagions, we show-
31 case monitoring and forecast tools for epidemic
32 crises centered around a crowd-sourced, real-time
33 method to assess levels of physical proximity in
34 a population using GPS location information, the
35 Contact Index CX [5]. We show that diverg-
36 ing trends between contact levels and independ-
37 ently recorded infection surveillance are indica-
38 tors of altered relative viral transmissibility. Using
39 2020-specific data as a baseline for purely contact-
40 driven SARS-CoV-2 epidemics, all observed transi-
41 tion points are explained by the onset of key im-
42 mune escape variants (alpha, delta, omicron). The
43 resulting dual evolution, Contact Index CX and re-
44 lative transmissibility T , provides a highly transpar-
45 ent and timely picture of ongoing epidemics, includ-
46 ing the possibility to identify likely driving forces in
47 future epidemic waves.

48 2. Materials and Methods

49 2.1. Contact metrics relevant for epidemics

50 Contact networks are a representation of hu-
51 man interactions [6] with immediate implications
52 for the spread of contagions in a population [7, 8]:
53 Nodes represent individuals and edges are drawn
54 between pairs of nodes in the event of contact be-
55 tween them (Figure 1(a,b)). A contagion can prop-

agate through a population along paths following
56 the links of the network.

57 Intuitively, transmission levels scale with the
58 average number of links per node $\langle k \rangle =$
59 $\sum_{k \geq 0} kP(k) = 2L/N$ [3], where $P(k)$ is the dis-
60 tribution of these numbers across a network and
61 N (L) is the number of nodes (links). Beyond
62 this local property, more global topological network
63 features – how contacts are collectively configured
64 across the network – do also affect the course of epi-
65 demics [3] by fueling and constraining the number
66 of available paths. Groundbreaking epidemiological
67 and network-theoretical work established that the
68 effective reproduction number R_{eff} , quantifying epi-
69 demic spreading, scales with $\frac{\langle k^2 \rangle}{\langle k \rangle}$ [3, 9, 10, 11, 12],
70 i.e. the presence of very social nodes (superspread-
71 ers) with outstanding k mediate enhanced propa-
72 gation. Typical social networks are very inhomoge-
73 neous in terms of social activity, with outstanding
74 community structure and few individuals responsi-
75 ble for most contacts [9]. The pivotal role of the
76 second moment $\langle k^2 \rangle = \sum_{k \geq 0} k^2 P(k)$ is intuited by
77 the *friendship paradox* [13]: An individual’s friends
78 are on average more social than oneself; in other
79 words, the number of next-nearest neighbors $\langle k^2 \rangle$ in
80 the network exceeds the expectation $\langle k \rangle^2$ from the
81 number of nearest neighbors, a mere consequence
82 of non-zero variance in $P(k)$: $\langle k^2 \rangle - \langle k \rangle^2 > 0$ (Supp
83 Mat S2).

85 2.2. Assessing contact levels in real-world networks

86 The contact network relevant to transmission of
87 airborne viruses such as Influenza and SARS-CoV-
88 2 arises from physical proximity between individ-
89 uals (Figure 1(a)). Compared to (virtual) social
90 networks, such real-world networks are expected to
91 have distinct properties, as they are constrained
92 by geography and physical distance, but are also
93 tremendously more difficult to track at the popula-
94 tion scale. Coarse contact and mixing patterns in
95 real-world networks have been inferred using lim-
96 ited data gathered from surveys [14, 15] or viral

97 phylogeny [16]. Locally confined real-world net- 137
98 works, such as on cruise ships [17], school cam- 138
99 campuses [18] or within towns [19] have been measured
100 using Bluetooth communication between nearby
101 mobile devices.

102 We use a previously developed approach to probe
103 population-scale real-world contact networks based
104 on crowd-sourced datasets of GPS locations [20, 5]
105 to measure the Contact Index $CX = \frac{\langle k^2 \rangle}{\langle k \rangle}$ as a
106 statistical measure of contact levels relevant for
107 epidemics [5]. The crowd-sourcing data is col-
108 lected in near real-time via opt-in from each of an
109 anonymized panel of 1 million mobile app users
110 (roughly 1% of Germany’s population) and con-
111 sists of ≈ 100 daily samples per device tagged with
112 time and GPS location information. It allows us
113 to reconstruct samples of the actual contact net-
114 work realized in the population: Contacts (links)
115 are drawn between devices (nodes) co-located in
116 space and time (Figure 1(a) and Supp Mat S1). Ex-
117 amples of reconstructed contact networks are shown
118 in Fig. 1(e).

119 2.3. Network sampling correction

120 The incomplete nature of such crowd-sourced
121 data represents a major challenge: Contacts from
122 uninvolved or inactive devices are not captured, giv-
123 ing rise to missing nodes and links in the network.
124 This aspect of our data can be crafted into a net-
125 work sampling framework [21, 22] in which nodes
126 and edges are randomly removed with probabili-
127 ties p and q , respectively (Figure 1(b,c) and Supp
128 Mat S3). p denotes the population share repre-
129 sented in the panel of app users, while q is inter-
130 preted as the rate f_{ij} of simultaneous samples from
131 pairs of app users (Figure 1(c)), a necessary con-
132 dition to detect a contact between users with indi-
133 vidual sample rates f_i and f_j , respectively. These
134 sampling parameters are subject to change over
135 time beyond daytime-related periodicity (see be-
136 low), mostly in response to software updates and

app usage (Figure 1(d)), and are heterogeneous in
138 space (Supp Mat S4 and Figure S3(a,b,c)).

139 For simplicity, we here use daily averages of sam-
140 ple rates. The rate f_{ij} of simultaneous samples
141 tends to exceed the expectation from individual fre-
142 quencies $f_i f_j$ under the hypothesis of independence
143 of distinct mobile devices, i.e. $f_{ij} > f_i f_j$, espe-
144 cially prior to February 2020 (Figure 1(d)); a ma-
145 jor app update in February 2020 has significantly
146 altered the daytime distribution and overall num-
147 ber of samples (Figure 1(d)). This apparent correla-
148 tion between devices stems from the non-uniformity
149 of the sampling activity over the day: Devices are
150 more active during daytime than at night, an effect
151 particularly prominent prior to February 2020 (Fig-
152 ure 1(d)). However, aside from a common daytime
153 pattern, devices show a predominantly independent
154 activity pattern from one another (Figure 1(d)):
155 At any given timepoint (2 min interval), squared
156 single-device distributions, i.e. $\frac{\rho_1(t)^2}{\int \rho_1(t)^2 dt}$, do cap-
157 ture the distribution of simultaneous samples $\rho_2(t)$
158 across the day well. Solely in consequence to the
159 daytime-related correlation, we are likely to slightly
160 underestimate the true value of q by using daily av-
161 erages.

162 Our improved mathematical modeling based
163 on Horvitz-Thompson theory disentangles actual
164 changes in contact levels from signals unrelated to
165 the users’ contact behaviour, including participa-
166 tion and activity levels in the user panel, but ex-
167 cluding correlation between devices, see above. We
168 thus achieve a persistent and comparable results
169 across the full time span since the beginning of mea-
170 surement in 2019 (Supp Mat S3 and Supp Mat S4).
171 In summary, we show that the Contact Index CX
172 of an unobserved complete network G can be re-
173 trieved from a network sample G^* obtained under
174 the described sampling scheme according to

$$CX - 1 = \frac{CX^* - 1}{pq_{\text{eff}}}, \quad (1)$$

175 where $CX^* = \frac{\langle k^{*2} \rangle^*}{\langle k^* \rangle^*}$ is the same quantity measured 214
176 within the network sample and q_{eff} is an effective 215
177 node sampling probability for networks of unique 216
178 contacts (see below). 217

179 Importantly, abstractions of contact networks ex- 218
180 ist in two distinct flavours: weighted versus un- 219
181 weighted [23]. Links may be endowed with weights 220
182 $w_{ij} \in \{0, 1, 2, \dots\}$ representing the duration or 221
183 multiplicity of contact between individuals i and 222
184 j [24] or simply indicate the presence or absence of 223
185 contact $a_{ij} = \text{sgn}(w_{ij}) \in \{0, 1\}$ (Figure 1(f)). In 224
186 the epidemiological context, we assume that net- 225
187 work topology, represented by a_{ij} , is more im- 226
188 portant than the recurrence of contacts between 227
189 the same individuals: For instance, the (statistical) 228
190 contribution to viral spread from a cluster of 229
191 short contacts at a crowded event would outpace a 230
192 lengthy contact between an isolated couple while in 231
193 lockdown. We thus focus on unweighted networks 232
194 and exclude contact duration in our analyses other 233
195 than in the fact that short contacts are unlikely to 234
196 be recorded during the random sampling inherent 235
197 to the crowd-sourcing method. 236

198 However, network sampling destroys topological 237
199 information about underlying complete networks 238
200 (Figure 1(f)); the success of Horvitz-Thompson the- 239
201 ory [21] to establish a connection between original 240
202 and sample networks relies in the use of weighted 241
203 links (Supp Mat S3). To establish the same con- 242
204 nection for unweighted networks, we devised a 243
205 Bayesian approach which identifies missing topo- 244
206 logical information as the weight distribution for 245
207 existing links in the complete network $P(w|w > 0)$ 246
208 and defines the edge sampling probability as 247

$$q_{\text{eff}} = P(w^* > 0|w > 0) = 1 - G_{w|w>0}(1 - q), \quad (2)$$

209 where $G_{w|w>0}(\xi) = \sum_{w>0} P(w|w > 0)\xi^w$ is the 249
210 probability generating function of $P(w|w > 0)$ 250
211 (Supp Mat S3). We find that available com- 251
212 plete real-world networks in various contexts [17, 252
213 18, 19] appear to show strikingly similar weight 253

distributions (Figure 1(g)), which suggests a uni-
versal shape of $P(w|w > 0)$ also applicable to
our problem. Here, “complete” refers to the
aspect that these networks represent a fraction
of the population ($p < 1$), but all contacts
within that sub-population are being detected
($q = 1$) – node sampling, but no edge sampling.
These distributions are consistent with power laws
 $P(w|w > 0) = w^{-(1+\alpha)}/\zeta(1 + \alpha)$ with small expo-
nents [25, 26] (Figure 1(g)), a repeatedly demon-
strated feature of complex networks [27] and be-
yond [28]. Yet, we do not imply that power laws
are the true mechanism behind network weights, as
a variety of other distribution classes are easily con-
founded with power laws [28, 29, 30], but merely use
it as a prior for $P(w|w > 0)$.

3. Results

3.1. Evolution of CX since 2019

By means of our refined correction method for
network sampling effects, we achieve a consistent
measurement of contact levels since the begin-
ning of crowd-sourcing in 2019, despite the time-
dependent sampling. That is, we cover the prelude
and entire course of the SARS-CoV-2 epidemic in
Germany (Figure 2(a)). The gap in February 2020
is explained by missing data due to the rollout of a
major crowd-sourcing software update.

Holiday season comes along with reduced CX un-
der normal conditions, as shown by the Fall and
Christmas breaks in 2019, thus showing a reduc-
tion of transmission-prone contacts. The onset of
the first SARS-CoV-2 wave in March 2020 induced
an unequivocally more pronounced drop in CX ,
probably explained by a more systematic cessation
of super-spreading activities. The dramatically al-
tered contact network structure during a lockdown
is depicted in Figure 1(e).

Since onset of the SARS-CoV-2 pandemic,
changes in contact behaviour as reflected by CX
underwent several periods of spiking (partial or

254 complete deregulation of mass events in fall 2020, 295
255 fall 2021 and spring 2022) and damping (winter 296
256 wave 2020, emergence of the omicron variant in 297
257 late 2021). Overall, a similar evolution is observed 298
258 between CX and the rigor of SARS-CoV-2-related 299
259 policy as measured by the Government-Response 300
260 Index [31] (Figure S1(a)), thus indicating broad 301
261 awareness of the situation at the population and 302
262 governance levels albeit no causal link shall be im- 303
263 plied. 304

264 Interestingly, recent CX values have not yet re- 305
265 turned to pre-pandemic levels by a factor of 2 to 3, 306
266 despite a return to no contact-related restrictions 307
267 in 2022. This suggests the existence of a hystere- 308
268 sis effect in addition to the fast response of CX 309
269 discussed above: The collective behaviour has not 310
270 returned to its unperturbed state in response to re- 311
271 laxed conditions, possibly as a result of continued 312
272 broad perception of disease risk [32, 33]. 313

273 From a dimensional viewpoint, CX represents 314
274 an average number of (next-nearest) contacts per 315
275 (nearest) contact: Comparing values of CX across 316
276 areas with vastly different population densities 317
277 within Germany supports our expectation that CX 318
278 scales (non-linearly) with the absolute propensity 319
279 of physical proximity between individuals (Fig-
280 ure S3(d) and Supp Mat S4).

281 3.2. Deciphering epidemic forces: contacts vs. rel- 320 282 ative transmissibility 321

283 In 2020, SARS-CoV-2 epidemic trends were pri- 322
284 marily driven by trends in contact levels, as both 323
285 immune escape variants and vaccines were not yet 324
286 relevant and relative SARS-CoV-2 transmissibility 325
287 – its intrinsic transmission probability per contact 326
288 – was thus constant (Figure 2(b)): Official daily 327
289 now-cast reproduction numbers R_{eff} , independently 328
290 recorded from national infection surveillance [34], 329
291 correlate well with daily CX , but CX shows a 330
292 time lead of approximately 2 – 3 weeks over R_{eff} 331
293 (Figure S1(a, right inset)) [5], explained by incu- 332
294 bation time as well as testing and reporting de- 333

lays. This underlines the predictive character of
real-time contact metrics for wild-type dominated
epidemics [20]. Since then, the correlation between
 R_{eff} and CX has repeatedly changed, with the re-
sulting signal quantifying shifts in relative transmis-
sibility accountable to key epidemic changes other
than contacts.

The effective reproduction number R_{eff} is defined
by $R_{\text{eff}} = \langle k \rangle \cdot U \cdot \tau$, where $\langle k \rangle$ denotes the contact
number per day, U the probability of transmission
per contact, and τ the mean duration of infectivity
in days. Both U and τ are determined by physio-
logical processes involved in transmission and, to-
gether, define the intrinsic transmission efficiency
(per contact) $T = U \cdot \tau$.

Furthermore, as we assume $CX = \frac{\langle k^2 \rangle}{\langle k \rangle}$
replaces $\langle k \rangle$, we replace the definition by
 $R_{\text{eff}} = (a + b \cdot CX) \cdot T$. A linear relationship of
this form between CX and R_{eff} is motivated by our
findings in 2020. We use values for a and b obtained
from a linear regression between CX and wild-type
 R_{eff} data at the optimal time delay of $\Delta t = 16$ days
(Figure S1(a, left inset) and Supp Mat S5). Upon
interpreting $R_{\text{WT}}(CX) \equiv a + b \cdot CX$ as the wild-type
specific reproduction number, we have that

$$R_{\text{eff}} = R_{\text{WT}}(CX) \cdot T, \quad (3)$$

where T represents relative transmissibility with re-
spect to wild-type in a fully susceptible population
($T_{\text{WT}} = 1$). Note that, in contrast to now-cast data,
Eq. (3) assigns reproduction numbers to the day of
contact/infection.

From independently recorded values for R_{eff} and
 CX , we can determine the relative transmissi-
bility of the contagion by factoring out contact-
related contributions from overall infection dynam-
ics as $T = \frac{R_{\text{eff}}}{R_{\text{WT}}(CX)}$ for any given day. We ex-
pect network-wide propagation of transmissibility-
related information to be slow compared to network
dynamics itself and, thus, T to undergo evolution
on longer timescales. We interpret fast signal in

334 T as random fluctuations from the measurement of
335 R_{eff} and capture actual trends by $\langle T \rangle$, centered av-
336 erages over sliding time windows of 2 months (Supp
337 Mat S5).

338 3.3. Epidemic evolution of relative SARS-CoV-2 339 transmissibility

340 The evolution of relative SARS-CoV-2 transmis-
341 sibility $\langle T \rangle$ is shown in Figure 2(b). This time se-
342 ries reenacts the various phases of the SARS-CoV-2
343 pandemic:

344 Relative SARS-CoV-2 transmissibility $\langle T \rangle$ is ap-
345 proximately equal to unity throughout 2020, an
346 initial period purely driven by unperturbed wild-
347 type epidemics that we used to “calibrate” CX and
348 R_{eff} which evolve on shorter timescales. It sub-
349 sequently follows a tug-of-war pattern shaped by
350 alternating epidemic forces beyond contacts: im-
351 mune escape variants and development of popula-
352 tion immunity through infection and vaccination.
353 Three waves of increased relative transmissibility
354 are explained by the takeover of fitter virus lin-
355 eages (Figure 2(b)), specifically alpha (spring 2021),
356 delta (summer 2021) and omicron BA.1/BA.2 (win-
357 ter 2021/22). We hypothesize that subsequent re-
358 laxation of $\langle T \rangle$ after each wave may be attributed to
359 natural immunity, while the superposed long-term
360 downward trend may be explained by the additional
361 immunity acquisition through (initial and booster)
362 vaccination campaigns. Interestingly, the effect of
363 omicron BA.4/BA.5 takeover in summer 2022 on
364 $\langle T \rangle$ is nowhere close to those of previous variants.

365 Comparing correlations with different parame-
366 ters rules out the possibility that the measured $\langle T \rangle$
367 is shaped by factors confounding the reproduction
368 numbers or CX values (Figure S1(b,c) and Supp
369 Mat S5). These possible confounders include viral
370 prevalence, CX itself through higher-order effects
371 from network sampling not captured by our mod-
372 eling and other topological network features (such
373 as clustering, small-world properties) as well as R_{eff}

374 itself through changes in testing strategies and sys-
375 tematic under-reporting of infections [35]. For in-
376 stance, testing individuals indiscriminately versus
377 focusing test capacities on suspected infection cases
378 may lead to incomparable snapshots of ongoing in-
379 fection dynamics. Overall, strong positive correla-
380 tion is exclusively observed between $\langle T \rangle$ and variant
381 dynamics (Figure S1(b,c)) [36]. In this analysis, we
382 use test positivity [37] and results from local preva-
383 lence studies [38] as proxies for overall prevalence.
384 Also, we neglect possible effects from network sam-
385 pling on different topological measures [39, 40], but
386 we expect trends to be conserved as long as the
387 sampling process remains unchanged.

388 We note the absence of seasonal oscillations in
389 $\langle T \rangle$ as well as clear signatures of mask mandates
390 (in effect across many social contexts between April
391 2020 and April 2022). A seasonal oscillation in $\langle T \rangle$,
392 larger values in winter and smaller values in sum-
393 mer, might be expected from the shift of human ac-
394 tivity between in- and outdoor settings. Also, pre-
395 vious research established the effectiveness of mask
396 usage at reducing transmission of respiratory dis-
397 eases (reviewed in [41]). Overall, our results sug-
398 gest that, at least in the epidemic stage of SARS-
399 CoV-2, infection rates were predominantly driven
400 by the strong variability in contacts as well as the
401 repeated emergence of more transmissible variants,
402 in line with previous findings [42, 43, 44].

403 3.4. Forecast of infection level and trend changes

404 The challenge of epidemic forecast consists in the
405 accurate prediction of current and future reproduc-
406 tion numbers R_{eff} . Using the rationale that trends
407 in infection levels carry the combined signature of
408 trends in contact and relative transmissibility lev-
409 els, we propose to construct predictions according
410 to

$$R_{\text{true}}(t) = R_{\text{WT}}(CX(t)) \cdot \langle T(t) \rangle, \quad (4)$$

411 where R_{true} is assigned to the projected day of con- 452
412 tact/infection. The key difference to Eq. (3) is the 453
413 use of $\langle T \rangle$ which eliminates noise from reproduc- 454
414 tion numbers. Importantly, we therefore expect 455
415 that our prediction R_{true} represents actual epidemic 456
416 trends (ground truth) more accurately than epi- 457
417 demic surveillance (R_{eff}). 458

418 Figure 3(a) shows R_{true} together with data from 459
419 infection surveillance, both plotted with respect to 460
420 their date of recording (assuming real-time CX 461
421 measurement). This shows how our prediction 462
422 overall anticipates current epidemic trends that 463
423 are observed via infection surveillance only about 464
424 $\Delta t = 2 - 3$ weeks later. Thus, we propose to use 465
425 our method as a tool for real-time infection surveil- 466
426 lance. 467

427 To extend forecasts beyond this horizon and pre- 468
428 dict future reproduction numbers, CX and $\langle T \rangle$ 469
429 themselves need to be projected beyond latest data. 470
430 For several choices of the current day t_0 , Figure 3(b) 471
431 showcases forecasts (R_{pred}) where CX and $\langle T \rangle$ are 472
432 continued beyond the last days of available data 473
433 (t_0 and $t_0 - \Delta t$, respectively) using autoregressive 474
434 integrated moving average (ARIMA) models prior 475
435 to applying Eq. (4) (Supp Mat S6). These fore- 476
436 casts outperform a null forecast based on a mere 477
437 ARIMA-type continuation of infection surveillance 478
438 data (R_{eff}), as shown by narrower distributions of 479
439 residuals ($R_{\text{pred}} - R_{\text{true}}$) across all choices of t_0 (Fig- 480
440 ure 3(b)). Furthermore, we highlight the broad ap- 481
441 plicability of our method to airborne infectious dis- 482
442 eases by performing an identical forecast analysis 483
443 for Influenza (Figure S2(a)), using coarser infection 484
444 surveillance data [45] and presuming a similar rela- 485
445 tionship between R_{eff} and CX as for SARS-CoV-2 486
446 (Supp Mat S6). 487

447 Most importantly, trend changes in epidemic 488
448 driving forces such as $\langle T \rangle$ and CX are indicators of 489
449 new phases in an epidemic. Timely detection of new 490
450 trends in these time series, e.g. using anomaly de- 491
451 tection methods, can provide valuable information

to estimate the risk of upcoming epidemic waves
and to predict their nature – whether dynamics is
fueled by contacts or increased transmission effi-
ciency. Such trend detection is potentially easier
to achieve but equally informative than the abil-
ity to accurately predict infection surveillance. The
onset of rising trends could shape decision-making
with regard to the effectiveness of health policies,
e.g. pharmaceutical and non-pharmaceutical in-
terventions for rising $\langle T \rangle$ and CX , respectively.
Figures 3(c) and S2(b) highlight rising and falling
trends in both CX and T for SARS-CoV-2 and In-
fluenza, respectively, akin to trends in stock prices.
For SARS-CoV-2, trend changes are timely indica-
tors of all major escape variant- and contact-driven
epidemic turning points (Figure 3(c)). Unlike for
SARS-CoV-2 in its epidemic stage, major upheavals
in relative transmissibility for Influenza are limited
to seasonality, with the notable exception of 2020,
presumably reflecting its endemic dynamics (Fig-
ure S2(b)).

473 Discussion

474 We presented a simple, yet insightful quantitative
475 method for a data-driven decomposition of overall
476 epidemic dynamics into contact-related and trans-
477 mission efficiency-related contributions. It relies
478 on both the availability of infection surveillance
479 data as well as crowd-sourced GPS location data
480 to detect and quantify physical proximity between
481 susceptible individuals. Its appeal resides in the
482 merely bivariate yet highly informative projection
483 of epidemics paving the way towards timely iden-
484 tification of driving forces in an ongoing epidemic
485 – human versus viral factors – and possibly effec-
486 tive mitigation strategies – pharmaceutical versus
non-pharmaceutical.

487 The approach can be used for epidemic forecast
488 in multiple ways. Recent and projected future val-
489 ues of CX and $\langle T \rangle$ can be used for short-term
490 ($2 - 3$ weeks) and long-term prediction of infection
491

492 or reproduction numbers, thus taking our previ- 533
493 ously described short-term forecast further [5]. Yet, 534
494 a timely detection of trend changes could reliably 535
495 forecast upcoming waves and their nature without 536
496 the necessity to accurately predict infection surveil- 537
497 lance data. These tools can lead towards a more 538
498 strategic approach to epidemic mitigation and po- 539
499 tentially save lives by reducing the spread of deadly 540
500 diseases.

501 Results from the presumably most systematically 542
502 tracked epidemic to date, SARS-CoV-2, draw the 543
503 picture of co-evolution within the virus-host rela- 544
504 tion: Increasing immunity levels in the host pop- 545
505 ulation alternate with step-wise adaptation of the 546
506 virus through immune-escape variants. Other fre- 547
507 quently discussed factors, including mask policies 548
508 and seasonality, are presumably still below the cur- 549
509 rent statistical resolution of our method, defined by 550
510 the sampling noise in the CX and R_{eff} time series.
511 Moreover, a larger impact of seasonal variation is
512 expected in the endemic phase of SARS-CoV-2 [46].

513 Our method is broadly applicable to airborne 554
514 contagions beyond SARS-CoV-2, but depends on 555
515 the availability of infection surveillance and crowd- 556
516 sourcing strategies that remain persistent over ex- 557
517 tended amounts of time. Changes in testing strat- 558
518 egy can lead to signal and biases unrelated to un- 559
519 derlying epidemic driving forces [35]. More cru- 560
520 cially, systematic infection surveillance is not im- 561
521 plemented beyond the case of SARS-CoV-2. We 562
522 illustrated a framework to correct for the effect 563
523 of varying sampling depth in the contact network. 564
524 Yet, higher-order effects in the signal can occur as 565
525 a result of sampling aspects not captured by our 566
526 mathematical modeling. In order to ensure valid 567
527 prognoses through our method, we advocate for sys- 568
528 tematic and persistent crowd-sourcing and infection 569
529 surveillance strategies across a variety of diseases 570
530 with epidemic potential.

531 Geographical resolution of our forecast method 571
532 is currently limited by the sampling depth, as the 572
573

estimation especially of higher moments of degree 533
distributions $P(k)$ becomes increasingly difficult as 534
smaller portions of the network are available. A 535
higher spatial resolution of contact and relative 536
transmissibility levels, with potential to locate the 537
origin of new variants of concern and define locally 538
targeted mitigation strategies, can be achieved by 539
e.g. increasing the panel of app users. 540

541 Our analysis assumes statics, but actual contact 542
networks are dynamic in nature [47, 48]: While 543
some contacts are frequently repeated (e.g. be- 544
tween household members), other contacts are ran- 545
domly redrawn on each occasion (e.g. in pub- 546
lic transportation), with implications for epidemic 547
spread [49, 50]. Our method can be improved by 548
analyzing contact data in light of existing models 549
of dynamic networks [51, 48].

550 Acknowledgment

551 This work was supported by grants from the 552
Federal Government of Germany through the 553
Federal Ministry for Economic Affairs and Cli- 554
mate Action (BMWK) for the project DAKI-FWS 555
(01MK21009A) and the Federal Ministry of Educa- 556
tion and Research (BMBF) for the project Optim- 557
Agent (031L0299).

558 References

- 559 [1] A. Rodríguez, H. Kamarthi, P. Agarwal, J. Ho, M. Pa- 560
tel, S. Sapre, B. A. Prakash, Data-centric epidemic 561
forecasting: A survey, arXiv preprint arXiv:2207.09370 562
(2022).
- 563 [2] A. Maxmen, Has covid taught us anything about pan- 564
demic preparedness?, *Nature* 596 (2021) 332–335.
- 565 [3] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, 566
A. Vespignani, Epidemic processes in complex 567
networks, *Rev. Mod. Phys.* 87 (2015) 925–979. 568
doi:10.1103/RevModPhys.87.925. 569
URL [https://link.aps.org/doi/10.1103/ 570
RevModPhys.87.925](https://link.aps.org/doi/10.1103/RevModPhys.87.925)
- 571 [4] T. Alamo, D. G. Reina, P. Millán Gata, V. M. Preciado, 572
G. Giordano, Data-driven methods for present and fu- 573
ture pandemics: Monitoring, modelling and managing,

- 574 Annual Reviews in Control 52 (2021) 448–464. doi: 626
575 <https://doi.org/10.1016/j.arcontrol.2021.05.003>. 627
576 URL [https://www.sciencedirect.com/science/ 628](https://www.sciencedirect.com/science/article/pii/S1367578821000419)
577 [article/pii/S1367578821000419](https://www.sciencedirect.com/science/article/pii/S1367578821000419) 629
- 578 [5] S. Rüdiger, S. Konigorski, A. Rakowski, J. A. Edelman, 630
579 D. Zernick, A. Thieme, C. Lippert, Predicting the sars- 631
580 cov-2 effective reproduction number using bulk contact 632
581 data from mobile phones, *Proceedings of the National 633*
582 *Academy of Sciences* 118 (31) (2021) e2026731118. 634
583 arXiv: [https://www.pnas.org/doi/pdf/10.1073/ 635](https://www.pnas.org/doi/pdf/10.1073/pnas.2026731118)
584 [pnas.2026731118](https://www.pnas.org/doi/pdf/10.1073/pnas.2026731118), doi:10.1073/pnas.2026731118. 636
585 URL [https://www.pnas.org/doi/abs/10.1073/pnas. 637](https://www.pnas.org/doi/abs/10.1073/pnas.2026731118)
586 [2026731118](https://www.pnas.org/doi/abs/10.1073/pnas.2026731118) 638
- 587 [6] M. E. J. Newman, J. Park, Why social networks are 639
588 different from other types of networks, *Phys. Rev. E* 68 640
589 (2003) 036122. doi:10.1103/PhysRevE.68.036122. 641
590 URL [https://link.aps.org/doi/10.1103/PhysRevE. 642](https://link.aps.org/doi/10.1103/PhysRevE.68.036122)
591 [68.036122](https://link.aps.org/doi/10.1103/PhysRevE.68.036122) 643
- 592 [7] M. Keeling, The implications of network structure 644
593 for epidemic dynamics, *Theoretical population biology* 645
594 67 (1) (2005) 1–8. 646
- 595 [8] C. Moore, M. E. J. Newman, Epidemics and percolation 647
596 in small-world networks, *Phys. Rev. E* 61 (2000) 5678– 648
597 5682. doi:10.1103/PhysRevE.61.5678. 649
598 URL [https://link.aps.org/doi/10.1103/PhysRevE. 650](https://link.aps.org/doi/10.1103/PhysRevE.61.5678)
599 [61.5678](https://link.aps.org/doi/10.1103/PhysRevE.61.5678) 651
- 600 [9] M. Barthélemy, A. Barrat, R. Pastor-Satorras, 652
601 A. Vespignani, Dynamical patterns of epidemic out- 653
602 breaks in complex heterogeneous networks, *Journal of 654*
603 *Theoretical Biology* 235 (2) (2005) 275–288. 655
604 doi:<https://doi.org/10.1016/j.jtbi.2005.01.011>. 656
605 URL [https://www.sciencedirect.com/science/ 657](https://www.sciencedirect.com/science/article/pii/S0022519305000251)
606 [article/pii/S0022519305000251](https://www.sciencedirect.com/science/article/pii/S0022519305000251) 658
- 607 [10] Y. Moreno, R. Pastor-Satorras, A. Vespignani, Epi- 659
608 demic outbreaks in complex heterogeneous networks, 660
609 *The European Physical Journal B-Condensed Matter 661*
610 *and Complex Systems* 26 (4) (2002) 521–529. 662
- 611 [11] R. M. May, R. M. Anderson, M. E. Irwin, R. M. 663
612 Anderson, J. M. Thresh, The transmission dynamics 664
613 of human immunodeficiency virus (hiv), *Philosophical 665*
614 *Transactions of the Royal Society of London. B, Bi- 666*
615 *ological Sciences* 321 (1207) (1988) 565–607. arXiv: 667
616 [https://royalsocietypublishing.org/doi/pdf/10. 668](https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.1988.0108)
617 [1098/rstb.1988.0108](https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.1988.0108), doi:10.1098/rstb.1988.0108. 669
618 URL [https://royalsocietypublishing.org/doi/abs/ 670](https://royalsocietypublishing.org/doi/abs/10.1098/rstb.1988.0108)
619 [10.1098/rstb.1988.0108](https://royalsocietypublishing.org/doi/abs/10.1098/rstb.1988.0108) 671
- 620 [12] R. M. May, R. M. Anderson, Transmission dynamics of 672
621 hiv infection, *Nature* 326 (1987) 137–142. 673
- 622 [13] S. L. Feld, Why your friends have more friends than 674
623 you do, *American Journal of Sociology* 96 (6) (1991) 675
624 1464–1477. arXiv:<https://doi.org/10.1086/229693>, 676
625 doi:10.1086/229693. 677
- URL <https://doi.org/10.1086/229693>
- [14] J. M. Read, K. T. Eames, W. J. Edmunds, Dynamic 678
social networks and the implications for the spread of 679
infectious disease, *Journal of The Royal Society Inter- 680*
face 5 (26) (2008) 1001–1007. 681
- [15] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Au- 682
ranen, R. Mikolajczyk, M. Massari, S. Salmaso, 683
G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska- 684
Todys, M. Rosinska, W. J. Edmunds, Social con- 685
tacts and mixing patterns relevant to the spread of 686
infectious diseases, *PLOS Medicine* 5 (3) (2008) 1–1. 687
doi:10.1371/journal.pmed.0050074. 688
URL [https://doi.org/10.1371/journal.pmed. 689](https://doi.org/10.1371/journal.pmed.0050074)
[0050074](https://doi.org/10.1371/journal.pmed.0050074) 690
- [16] G. E. Leventhal, R. Kouyos, T. Stadler, V. von 691
Wyl, S. Yerly, J. Böni, C. Celleraï, T. Klimkait, 692
H. F. Günthard, S. Bonhoeffer, Inferring epi- 693
demic contact structure from phylogenetic trees, 694
PLOS Computational Biology 8 (3) (2012) 1–10. 695
doi:10.1371/journal.pcbi.1002413. 696
URL [https://doi.org/10.1371/journal.pcbi. 697](https://doi.org/10.1371/journal.pcbi.1002413)
[1002413](https://doi.org/10.1371/journal.pcbi.1002413) 698
- [17] R. Pung, J. A. Firth, L. G. Spurgin, V. J. Lee, A. J. 699
Kucharski, Using high-resolution contact networks to 700
evaluate sars-cov-2 transmission and control in large- 701
scale multi-day events, *Nature communications* 13 (1) 702
(2022) 1–11. 703
- [18] P. Sapiezynski, A. Stopczynski, D. D. Lassen, 704
S. Lehmann, Interaction data from the copenhagen net- 705
works study, *Scientific Data* 6 (1) (2019) 315. 706
- [19] S. M. Kissler, P. Klepac, M. Tang, A. J. Conlan, J. R. 707
Gog, Sparking “the bbc four pandemic”: Leveraging 708
citizen science and mobile phones to model the spread 709
of disease, *bioRxiv* (2020). arXiv:[https://www.biorxiv. 710](https://www.biorxiv.org/content/early/2020/05/12/479154.full.pdf)
[org/content/early/2020/05/12/479154.full.pdf](https://www.biorxiv.org/content/early/2020/05/12/479154.full.pdf), 711
doi:10.1101/479154. 712
URL [https://www.biorxiv.org/content/early/2020/ 713](https://www.biorxiv.org/content/early/2020/05/12/479154)
[05/12/479154](https://www.biorxiv.org/content/early/2020/05/12/479154) 714
- [20] F. W. Crawford, S. A. Jones, M. Cartter, S. G. 715
Dean, J. L. Warren, Z. R. Li, J. Barbieri, J. Camp- 716
bell, P. Kenney, T. Valleau, O. Morozova, Impact 717
of close interpersonal contact on covid-19 incidence: 718
Evidence from 1 year of mobile device data, *Science 719*
Advances 8 (1) (2022) eabi5499. arXiv:[https:// 720](https://www.science.org/doi/pdf/10.1126/sciadv.abi5499)
www.science.org/doi/pdf/10.1126/sciadv.abi5499, 721
doi:10.1126/sciadv.abi5499. 722
URL [https://www.science.org/doi/abs/10.1126/ 723](https://www.science.org/doi/abs/10.1126/sciadv.abi5499)
[sciadv.abi5499](https://www.science.org/doi/abs/10.1126/sciadv.abi5499) 724
- [21] E. D. Kolaczyk, *Statistical Analysis of Network Data*, 725
Springer New York, NY, 2009. doi:[https://doi.org/ 726](https://doi.org/10.1007/978-0-387-88146-1)
[10.1007/978-0-387-88146-1](https://doi.org/10.1007/978-0-387-88146-1). 727
- [22] P. Hu, W. C. Lau, A survey and taxonomy of graph 728

- 678 sampling, CoRR abs/1308.5865 (2013). arXiv:1308. 730
679 5865. 731
680 URL <http://arxiv.org/abs/1308.5865> 732
- [23] M. A. Serrano, M. Boguñá, R. Pastor-Satorras, Correla- 733
682 tions in weighted networks, Phys. Rev. E 74 (2006) 734
683 055101. doi:10.1103/PhysRevE.74.055101. 735
684 URL [https://link.aps.org/doi/10.1103/PhysRevE. 736
685 74.055101](https://link.aps.org/doi/10.1103/PhysRevE.74.055101) 737
- [24] A. Barrat, M. Barthélemy, R. Pastor-Satorras, 738
687 A. Vespignani, The architecture of complex weighted 739
688 networks, Proceedings of the National Academy of 740
689 Sciences 101 (11) (2004) 3747–3752. arXiv:<https://www.pnas.org/doi/pdf/10.1073/pnas.0400087101>, 741
690 doi:10.1073/pnas.0400087101. 742
691 URL [https://www.pnas.org/doi/abs/10.1073/pnas. 743
692 0400087101](https://www.pnas.org/doi/abs/10.1073/pnas.0400087101) 744
- [25] H. Ebel, L.-I. Mielsch, S. Bornholdt, Scale-free topology 745
694 of e-mail networks, Phys. Rev. E 66 (2002) 035103. doi: 746
695 10.1103/PhysRevE.66.035103. 747
696 URL [https://link.aps.org/doi/10.1103/PhysRevE. 748
697 66.035103](https://link.aps.org/doi/10.1103/PhysRevE.66.035103) 749
- [26] M. E. J. Newman, Scientific collaboration networks. i. 750
699 network construction and fundamental results, Phys. 751
700 Rev. E 64 (2001) 016131. doi:10.1103/PhysRevE.64. 752
701 016131. 753
702 URL [https://link.aps.org/doi/10.1103/PhysRevE. 754
703 64.016131](https://link.aps.org/doi/10.1103/PhysRevE.64.016131) 755
- [27] R. Guimera, S. Mossa, A. Turtschi, L. N. Amaral, 756
704 The worldwide air transportation network: Anoma- 757
705 lous centrality, community structure, and cities' global 758
706 roles, Proceedings of the National Academy of Sciences 759
707 102 (22) (2005) 7794–7799. 760
- [28] A. Clauset, C. R. Shalizi, M. E. J. Newman, Power-law 761
710 distributions in empirical data, SIAM Review 51 (4) 762
711 (2009) 661–703. arXiv:[https://doi.org/10.1137/ 763
712 070710111](https://doi.org/10.1137/070710111), doi:10.1137/070710111. 764
713 URL <https://doi.org/10.1137/070710111> 765
- [29] J.-D. J. Han, D. Dupuy, N. Bertin, M. E. Cusick, 766
714 M. Vidal, Effect of sampling on topology predictions 767
715 of protein-protein interaction networks, Nature biotech- 768
716 nology 23 (7) (2005) 839–844. 769
717
- [30] R. Perline, Strong, weak and false inverse power laws, 770
718 Statistical Science 20 (1) (2005) 68–88. 771
719 URL <http://www.jstor.org/stable/20061161> 772
- [31] T. Hale, N. Angrist, R. Goldszmidt, B. Kira, A. Pether- 773
722 ick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hal- 774
723 las, S. Majumdar, et al., A global panel database of pan- 775
724 demic policies (oxford covid-19 government response 776
725 tracker), Nature human behaviour 5 (4) (2021) 529– 777
726 538. 778
727
- [32] C. Betsch, S. Eitze, P. Sprengholz, L. Korn, P. Sham- 780
728 srizi, M. Geiger, E. Sievert, L. Lehrer, M. Jenny, Zusam- 781
729 menfassung und empfehlungen welle 69 (2022).
URL [https://projekte.uni-erfurt.de/cosmo2020/
web/summary/69/](https://projekte.uni-erfurt.de/cosmo2020/web/summary/69/)
- [33] C. Betsch, L. Wieler, M. Bosnjak, M. Ramharter, 782
733 V. Stollorz, S. Omer, L. Korn, P. Sprengholz, L. Fel- 734
735 gendreff, S. Eitze, P. Schmid, Germany covid-19 snap- 736
737 shot monitoring (cosmo germany): Monitoring knowl- 738
739 edge, risk perceptions, preventive behaviours, and pub- 740
741 lic trust in the current coronavirus outbreak in germany 742
(Mar. 2020). doi:10.23668/psycharchives.2776.
URL [https://psycharchives.org/index.php/en/
item/e5acdc65-77e9-4fd4-9cd2-bf6aa2dd5eba](https://psycharchives.org/index.php/en/item/e5acdc65-77e9-4fd4-9cd2-bf6aa2dd5eba) 743
- [34] M. an der Heiden, Sars-cov-2-nowcasting und -r- 744
745 schatzung (Jan. 2023). doi:10.5281/zenodo.7571376.
URL <https://doi.org/10.5281/zenodo.7571376> 746
- [35] H. Rossman, E. Segal, Nowcasting the spread of sars- 747
748 cov-2, Nature microbiology 7 (1) (2022) 16–17. 749
- [36] R. Koch-Institut, Anzahl und anteile von voc und voi 750
751 in deutschland (2023).
URL [https://www.rki.de/DE/Content/InfAZ/N/
Neuartiges_Coronavirus/Daten/VOC_VOI_Tabelle.
xlsx?__blob=publicationFile](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/VOC_VOI_Tabelle.xlsx?__blob=publicationFile) 752
- [37] E. Mathieu, H. Ritchie, L. Rodés-Guirao, C. Appel, 753
754 C. Giattino, J. Hasell, B. Macdonald, S. Dattani, 755
756 D. Beltekian, E. Ortiz-Ospina, M. Roser, Coronavirus 757
758 pandemic (covid-19) (2020).
URL <https://ourworldindata.org/coronavirus> 759
- [38] H. Neuhauser, N. Buttman-Schweiger, J. Fiebig, 760
761 C. Poethko-Müller, F. Prütz, G. Sarganas Margolis, 762
763 R. Thamm, M. Zimmermann, Observatorium serolo- 764
765 gischer Studien zu SARS-CoV-2 in Deutschland (Sep. 766
767 2022). doi:10.5281/zenodo.7043025.
URL <https://doi.org/10.5281/zenodo.7043025> 768
- [39] J. D. Noh, Percolation transition in networks with 769
770 degree-degree correlation, Phys. Rev. E 76 (2007) 771
772 026116. doi:10.1103/PhysRevE.76.026116.
URL [https://link.aps.org/doi/10.1103/PhysRevE. 773
774 76.026116](https://link.aps.org/doi/10.1103/PhysRevE.76.026116) 775
- [40] S. H. Lee, P.-J. Kim, H. Jeong, Statistical properties 776
777 of sampled networks, Phys. Rev. E 73 (2006) 016102.
doi:10.1103/PhysRevE.73.016102.
URL [https://link.aps.org/doi/10.1103/PhysRevE. 778
779 73.016102](https://link.aps.org/doi/10.1103/PhysRevE.73.016102) 780
- [41] J. T. Brooks, J. C. Butler, Effectiveness of 781
782 Mask Wearing to Control Community Spread 783
784 of SARS-CoV-2, JAMA 325 (10) (2021) 998– 785
786 999. arXiv:[https://jamanetwork.com/journals/
jama/articlepdf/2776536/jama_1_brooks_ 787
788 _2021_it_210006_1631033869_97869.pdf](https://jamanetwork.com/journals/jama/articlepdf/2776536/jama_1_brooks_2021_it_210006_1631033869_97869.pdf),
doi:10.1001/jama.2021.1505.
URL <https://doi.org/10.1001/jama.2021.1505> 789
- [42] F. Balloux, C. Tan, L. Swadling, D. Richard, C. Jenner, 790

- 782 M. Maini, L. van Dorp, The past, current and future 834
783 epidemiological dynamic of SARS-CoV-2, Oxford Open 835
784 Immunology 3 (1) (06 2022). arXiv:<https://academic.oup.com/oiim/article-pdf/3/1/iqac003/48744431/iqac003.pdf>, doi:10.1093/oxfimm/iqac003. 836
785 URL <https://doi.org/10.1093/oxfimm/iqac003> 837
786 [43] G. P. Guy Jr, F. C. Lee, G. Sunshine, R. McCord, 838
787 M. Howard-Williams, L. Kompaniyets, C. Dunphy, 839
788 M. Gakh, R. Weber, E. Sauber-Schatz, et al., Asso- 840
789 ciation of state-issued mask mandates and allowing on- 841
790 premises restaurant dining with county-level covid-19 842
791 case and death growth rates—united states, march 1– 843
792 december 31, 2020, Morbidity and Mortality Weekly 844
793 Report 70 (10) (2021) 350. 845
794 [44] D. K. Chu, E. A. Akl, S. Duda, K. Solo, S. Yaacoub, 846
795 H. J. Schünemann, D. K. Chu, E. A. Akl, A. El-harakeh, 847
796 A. Bognanni, T. Lotfi, M. Loeb, A. Hajizadeh, A. Bak, 848
797 A. Izcovich, C. A. Cuello-Garcia, C. Chen, D. J. Har- 849
798 ris, E. Borowiack, F. Chamseddine, F. Schünemann, 850
799 G. P. Morgano, G. E. U. Muti Schünemann, G. Chen, 851
800 H. Zhao, I. Neumann, J. Chan, J. Khabsa, L. Hneiny, 852
801 L. Harrison, M. Smith, N. Rizk, P. Giorgi Rossi, P. Abi- 853
802 Hanna, R. El-khoury, R. Stalteri, T. Baldeh, T. Piggott, 854
803 Y. Zhang, Z. Saad, A. Khamis, M. Reinap, S. Duda, 855
804 K. Solo, S. Yaacoub, H. J. Schünemann, Physical dis- 856
805 tancing, face masks, and eye protection to prevent 857
806 person-to-person transmission of sars-cov-2 and covid- 858
807 19: a systematic review and meta-analysis, The Lancet 859
808 395 (10242) (2020) 1973–1987. doi:[https://doi.org/10.1016/S0140-6736\(20\)31142-9](https://doi.org/10.1016/S0140-6736(20)31142-9). 860
809 URL <https://www.sciencedirect.com/science/article/pii/S0140673620311429>
810 [45] R. Koch-Institut, *Survstat@rki* 2.0 (2023).
811 URL <https://survstat.rki.de/>
812 [46] J. P. Townsend, A. D. Lamb, H. B. Hassler, 813
814 P. Sah, A. A. Nishio, C. Nguyen, A. D. Tew, 815
816 A. P. Galvani, A. Dornburg, Projecting the sea- 817
818 sonality of endemic covid-19, medRxiv (2022). 818
819 arXiv: <https://www.medrxiv.org/content/early/2022/10/07/2022.01.26.22269905.full.pdf>, doi:10. 819
820 1101/2022.01.26.22269905. 820
821 URL <https://www.medrxiv.org/content/early/2022/10/07/2022.01.26.22269905>
822 [47] V. Sekara, A. Stopczynski, S. Lehmann, Fun- 823
824 damental structures of dynamic social networks, 824
825 Proceedings of the National Academy of Sciences 825
826 113 (36) (2016) 9977–9982. arXiv: <https://www.pnas.org/doi/pdf/10.1073/pnas.1602803113>, 826
827 doi:10.1073/pnas.1602803113. 827
828 URL <https://www.pnas.org/doi/abs/10.1073/pnas.1602803113>
829 [48] P. Holme, J. Saramäki, Temporal networks, Physics Re- 829
830 ports 519 (3) (2012) 97–125, temporal Networks. doi: 830
831 <https://doi.org/10.1016/j.physrep.2012.03.001>. 831
832 URL <https://www.sciencedirect.com/science/article/pii/S0370157312000841>
833 [49] J. Enright, R. R. Kao, Epidemics on dynamic networks, 833
834 Epidemics 24 (2018) 88–97. doi:<https://doi.org/10.1016/j.epidem.2018.04.003>. 834
835 URL <https://www.sciencedirect.com/science/article/pii/S1755436518300173>
836 [50] E. Valdano, L. Ferreri, C. Poletto, V. Colizza, An- 836
837 alytical computation of the epidemic threshold on 837
838 temporal networks, Phys. Rev. X 5 (2015) 021005. 838
839 doi:10.1103/PhysRevX.5.021005. 839
840 URL <https://link.aps.org/doi/10.1103/PhysRevX.5.021005>
841 [51] X. Zhang, C. Moore, M. E. Newman, Random graph 841
842 models for dynamic networks, The European Physical 842
843 Journal B 90 (10) (2017) 1–14. 843
844 [52] N. Pitsianis, D. Floros, A.-S. Iliopoulos, X. Sun, Sg- 844
845 t-sne- π : Swift neighbor embedding of sparse stochastic 845
846 graphs, Journal of Open Source Software 4 (39) (2019) 846
847 1577. 847
848 [53] N. Pitsianis, A.-S. Iliopoulos, D. Floros, X. Sun, Space- 848
849 land embedding of sparse stochastic graphs, in: 2019 849
850 IEEE High Performance Extreme Computing Confer- 850
851 ence (HPEC), 2019, pp. 1–8. doi:10.1109/HPEC.2019. 851
852 8916505. 852

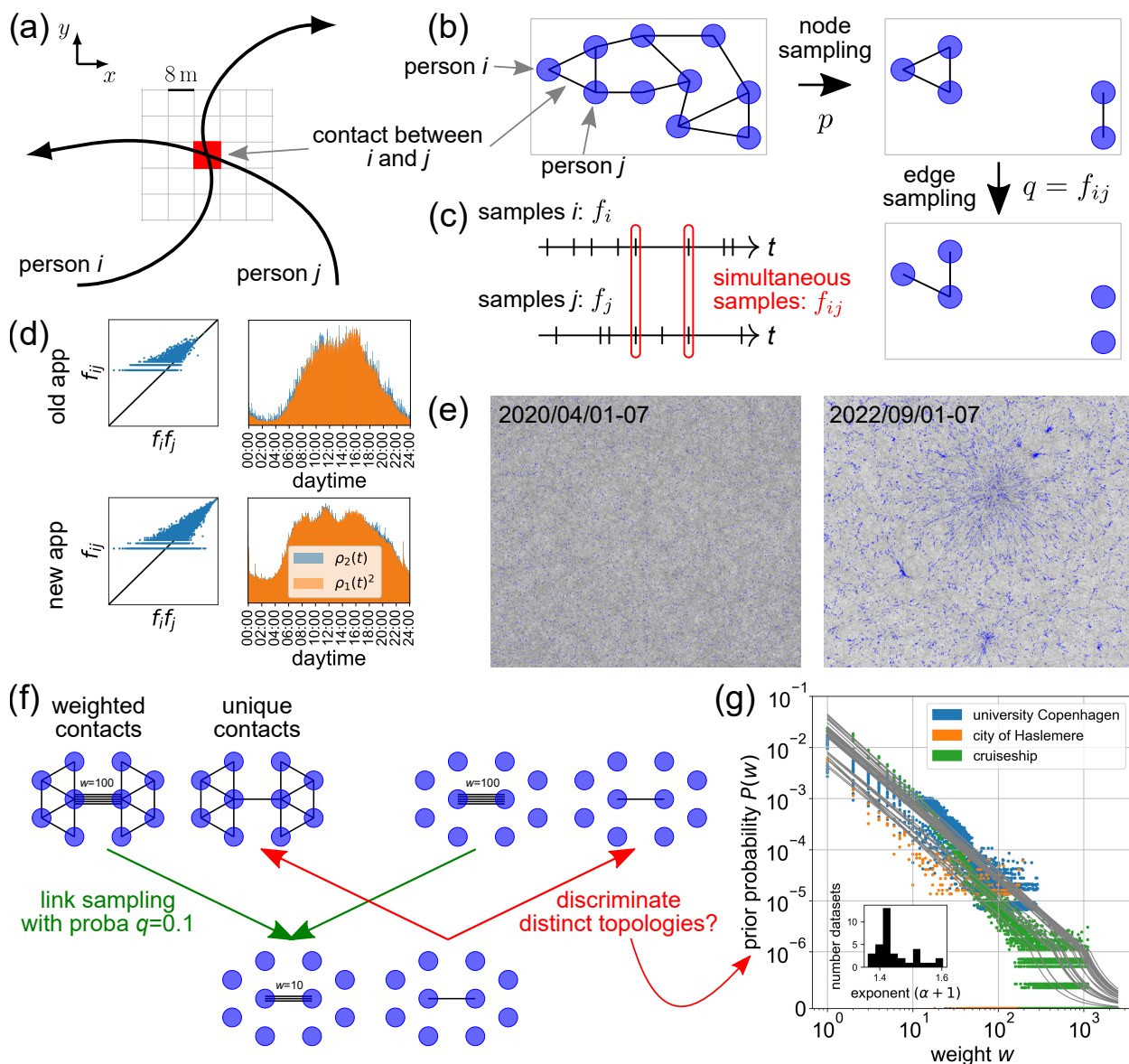


Figure 1: **Contact networks: definition, measurement and inference.** (a) Co-location of 2 (or more) devices i and j within the same $8\text{ m} \times 8\text{ m}$ cell within 2 min defines a contact. (b) In the network, pairs of individuals/mobile devices (nodes) are connected by their contacts (edges). The network sampling induced by the data collection app retains nodes with proba p (incl. links between pairs of retained nodes), reflective of the population share of app users. Subsequently, links are retained with proba q . (c) Ticks along the time axes indicate samples from a pair of devices i and j . q depends on the likelihood f_{ij} of simultaneous samples (red encircled samples), a necessary condition to observe a contact between them. (d) Comparison between actual simultaneous sample rates f_{ij} and those predicted from uncorrelated single-device sample rates $f_i f_j$ (left panel) and between the distribution of simultaneous samples over the day $\rho_2(t)$ with the squared distribution of single-device samples $\rho_1(t)^2$ (right panel). (e) Examples of 7-day aggregated networks under lockdown (Apr 2020) and unrestricted (Sep 2022) conditions. Blue dots represent individuals, gray links the contacts. Zoom over a 2D embedding using SG-t-SNE-II [52, 53]. (f) In weighted contact networks, links are weighted by the duration/multiplicity of contact $w_{ij} \in \{0, 1, 2, \dots\}$ between nodes i and j , while unique contact networks only distinguish between presence or absence of contact, $a_{ij} \in \{0, 1\}$. Example of information loss upon link sampling: Networks with distinct topologies (left vs. right set of networks) can yield similar sample networks (bottom networks) upon the same sampling process (green arrows). Discriminating distinct original networks from the sample network (red arrows) thus requires additional information. (g) Prior information is extracted from weight distributions $P(w)$ found in complete contact networks [17, 18, 19].

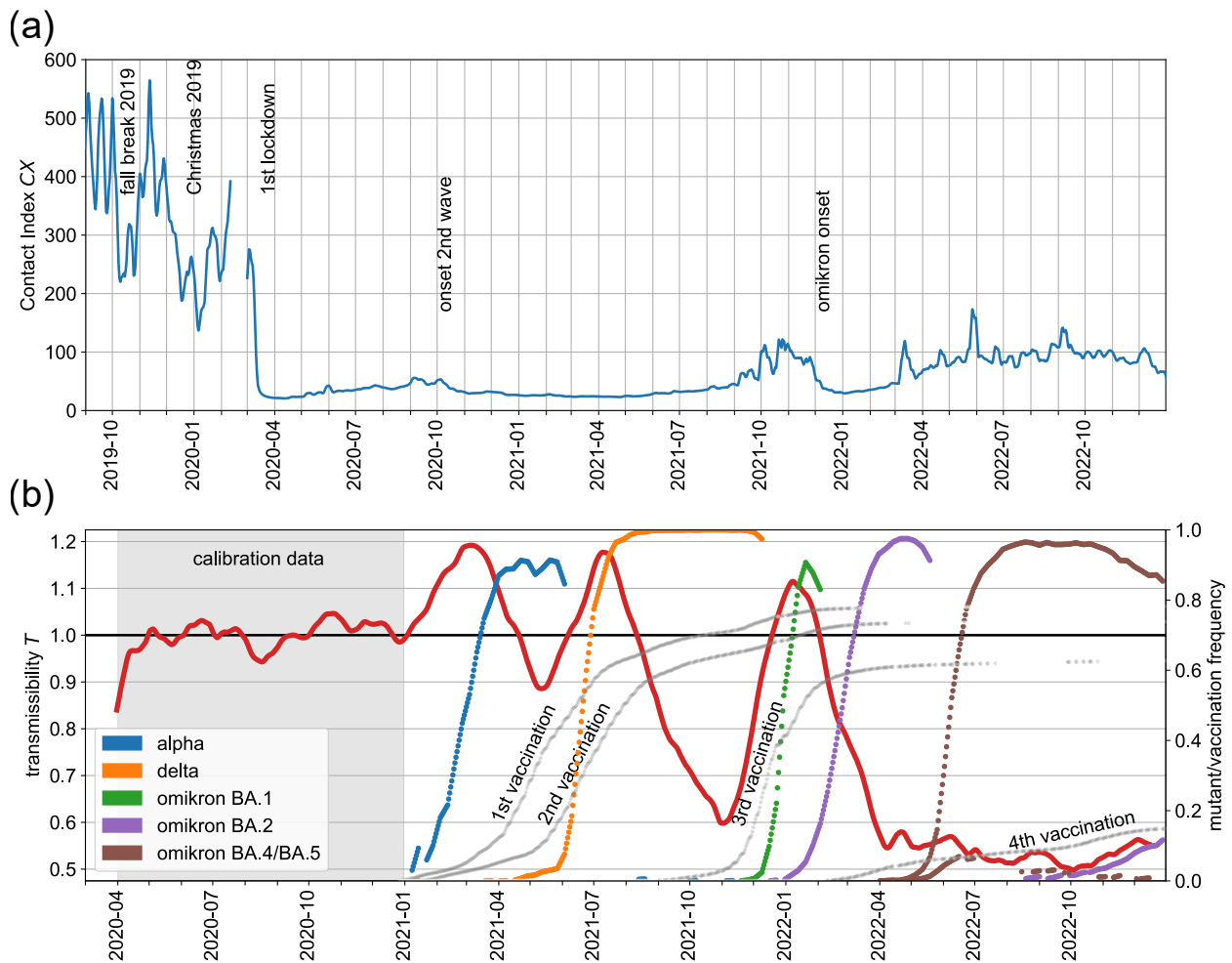


Figure 2: Real-time observation of driving forces in SARS-CoV-2 epidemics: contact levels and relative transmissibility. (a) Evolution of the Contact Index $CX = \langle k^2 \rangle / \langle k \rangle$ in Germany over the course of > 3 years (2019-2022), carrying the signature of various collective behaviour changes in response to the epidemic situation (as indicated). The gap in February 2022 is explained by a major app update. (b) The slowly varying relative transmissibility $\langle T \rangle(t)$ (red) quantifying the intrinsic efficiency of SARS-CoV-2 transmission, measured from the ratio of reproduction numbers (R_{eff}) and contact levels (CX), see Eq. (3). The gray-shaded time interval is wild-type dominated and was used to calibrate CX from our crowd-sourcing method and R_{eff} from infection surveillance (Figure S1(a, inset)). The rising frequencies of key SARS-CoV-2 immune escape variants (colored lines, see legend) and well as of vaccine status in Germany (light gray lines) are shown (right axis).

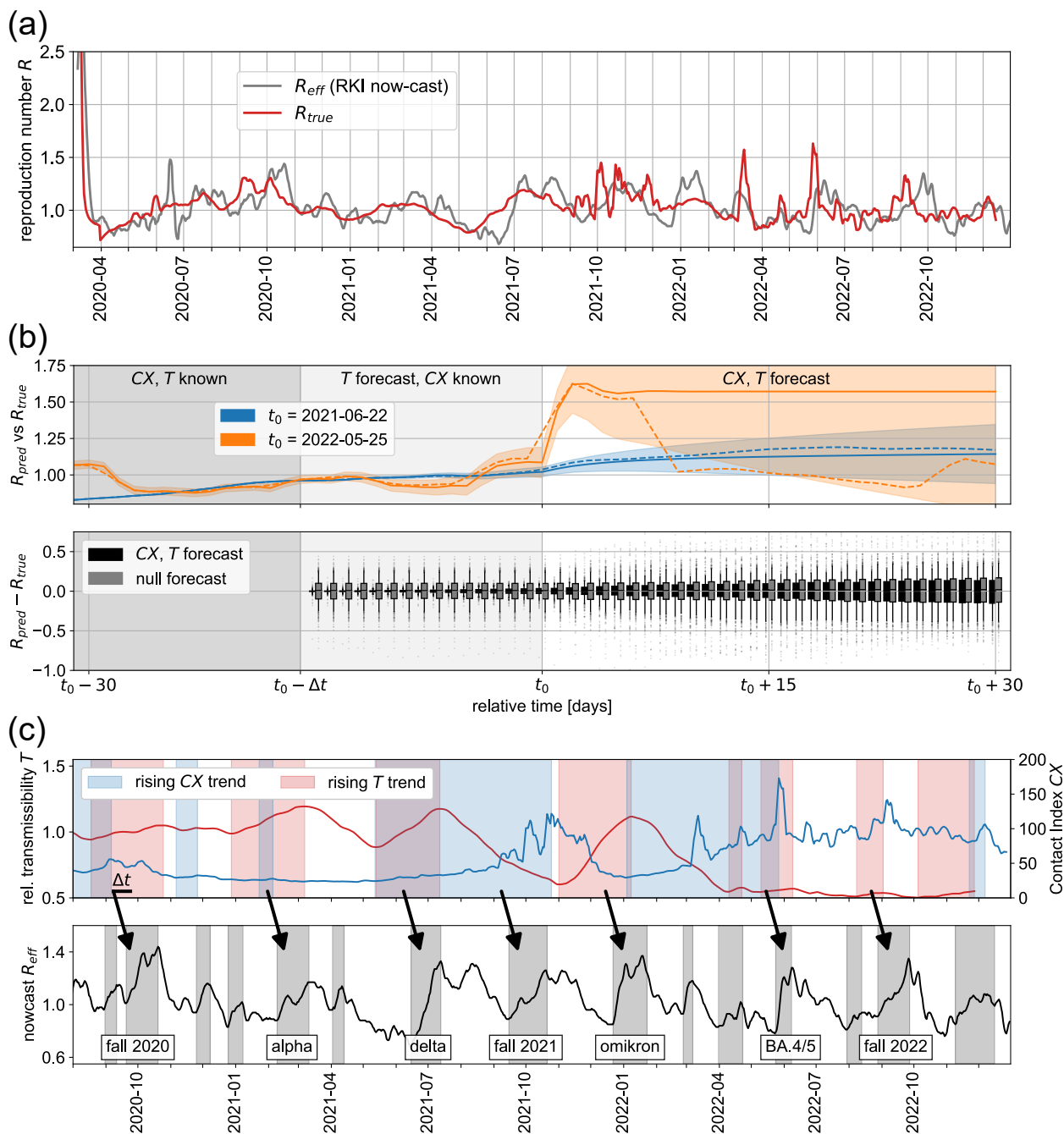


Figure 3: Forecast of reproduction numbers and trends from contact and transmissibility levels. (a) Comparison of SARS-CoV-2 effective reproduction numbers R_{eff} from infection surveillance (gray) and projected R_{true} using Eq. (4) (red). All reproduction numbers are assigned to their day of recording. **(b, upper panel)** Forecast R_{pred} of current and future SARS-CoV-2 reproduction numbers and their uncertainties (solid lines and shaded bands, respectively) using Eq. (4) and the CX and $\langle T \rangle$ time series. Comparison with actual R_{true} values (dashed lines). Denoting the current day by t_0 , R_{eff} and $\langle T \rangle$ are available up to $t_0 - \Delta t$, while CX is near real-time (available up to t_0); the time series are projected beyond their last time points using ARIMA models. The forecast is shown for different choices of the current day t_0 (see legend). **(b, lower panel)** The distribution of residuals between forecasted R_{pred} and actual R_{true} values over all choices of t_0 over the course of 2 years (black box plots). Comparison to residuals from null projections of R_{eff} that make no use of CX (gray box plots), i.e. simple ARIMA model-based projection of infection surveillance data. The boxes indicate quartiles, while whiskers cover 90% of the data. **(c)** Identification of rising trends in both contact levels and transmission efficiency (upper panel) and their relation to rising trends in R_{eff} (lower panel).