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

^aMachine Learning and Health Unit, Department of Engineering, NET CHECK GmbH, Berlin, Germany

^bDepartment of Biostatistics, Yale School of Public Health, New Haven, CT, USA

^cDepartment of Statistics and Data Science, Yale University, New Haven, CT, USA

^dDepartment of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

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

^fInstitute 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 1. Introduction

Infectious diseases represent serious threats to
an ever increasingly connected humankind, on par
with e.g. natural disasters and infrastructure failures. Epidemic preparedness – the ability to predict and mitigate future epidemic outbreaks – has
thus risen to one of the most pressing challenges in

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

Epidemic dynamics are shaped at the crossroads of human and viral driving forces: a pathogen's reproductive cycle, defining its relative transmission rate upon physical proximity between individuals with full or partial susceptibility, as well as hu-2023-07-25

15

16

^{*}Corresponding author: steven.schulz@netcheck.de.

56

57

50

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

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

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

⁴⁸ 2. Materials and Methods

49 2.1. Contact metrics relevant for epidemics

⁵⁰ Contact networks are a representation of hu⁵¹ 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 be⁵⁵ tween them (Figure 1(a,b)). A contagion can prop-

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

Intuitively, transmission levels scale with the average number of links per node $\langle k \rangle$ = $\sum_{k>0} kP(k) = 2L/N$ [3], where P(k) is the distribution of these numbers across a network and N (L) is the number of nodes (links). Beyond this local property, more global topological network features – how contacts are collectively configured across the network – do also affect the course of epidemics [3] by fueling and constraining the number of available paths. Groundbreaking epidemiological and network-theoretical work established that the effective reproduction number $R_{\rm eff}$, quantifying epidemic spreading, scales with $\frac{\langle k^2 \rangle}{\langle k \rangle}$ [3, 9, 10, 11, 12], i.e. the presence of very social nodes (superspreaders) with outstanding k mediate enhanced propagation. Typical social networks are very inhomogeneous in terms of social activity, with outstanding community structure and few individuals responsible for most contacts [9]. The pivotal role of the second moment $\langle k^2 \rangle = \sum_{k \ge 0} k^2 P(k)$ is intuited by the *friendship paradox* [13]: An individual's friends are on average more social than oneself; in other words, the number of next-nearest neighbors $\langle k^2 \rangle$ in the network exceeds the expectation $\langle k \rangle^2$ from the number of nearest neighbors, a mere consequence of non-zero variance in P(k): $\langle k^2 \rangle - \langle k \rangle^2 > 0$ (Supp Mat S2).

2.2. Assessing contact levels in real-world networks

The contact network relevant to transmission of airborne viruses such as Influenza and SARS-CoV-2 arises from physical proximity between individuals (Figure 1(a)). Compared to (virtual) social networks, such real-world networks are expected to have distinct properties, as they are constrained by geography and physical distance, but are also tremendously more difficult to track at the population scale. Coarse contact and mixing patterns in real-world networks have been inferred using limited data gathered from surveys [14, 15] or viral

141

142

144

147

148

149

150

151

154

155

156

157

158

159

160

161

163

164

165

166

167

168

171

phylogeny [16]. Locally confined real-world net- 137 97 works, such as on cruise ships [17], school cam-138 puses [18] or within towns [19] have been measured 99 139 using Bluetooth communication between nearby 100 140 mobile devices. 101

We use a previously developed approach to probe 102 population-scale real-world contact networks based 103 143 on crowd-sourced datasets of GPS locations [20, 5] 104 to measure the Contact Index $CX = \frac{\langle k^2 \rangle}{\langle k \rangle}$ as a 105 145 statistical measure of contact levels relevant for 106 146 epidemics [5]. The crowd-sourcing data is col-107 lected in near real-time via opt-in from each of an 108 anonymized panel of 1 million mobile app users 109 (roughly 1% of Germany's population) and con-110 sists of ≈ 100 daily samples per device tagged with 111 time and GPS location information. It allows us 112 152 to reconstruct samples of the actual contact net-113 153 work realized in the population: Contacts (links) 114 are drawn between devices (nodes) co-located in 115 space and time (Figure 1(a) and Supp Mat S1). Ex-116 amples of reconstructed contact networks are shown 117 in Fig. 1(e). 118

2.3. Network sampling correction 119

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

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

For simplicity, we here use daily averages of sample rates. The rate f_{ij} of simultaneous samples tends to exceed the expectation from individual frequencies $f_i f_j$ under the hypothesis of independence of distinct mobile devices, i.e. $f_{ij} > f_i f_j$, especially prior to February 2020 (Figure 1(d)); a major app update in February 2020 has significantly altered the daytime distribution and overall number of samples (Figure 1(d)). This apparent correlation between devices stems from the non-uniformity of the sampling activity over the day: Devices are more active during daytime than at night, an effect particularly prominent prior to February 2020 (Figure 1(d)). However, aside from a common daytime pattern, devices show a predominantly independent activity pattern from one another (Figure 1(d)): At any given timepoint $(2 \min \text{ interval})$, squared single-device distributions, i.e. $\frac{\rho_1(t)^2}{\int \rho_1(t)^2 dt}$, do capture the distribution of simultaneous samples $\rho_2(t)$ across the day well. Solely in consequence to the daytime-related correlation, we are likely to slightly underestimate the true value of q by using daily averages.

Our improved mathematical modeling based on Horvitz-Thompson theory disentangles actual changes in contact levels from signals unrelated to the users' contact behaviour, including participation and activity levels in the user panel, but excluding correlation between devices, see above. We thus achieve a persistent and comparable results across the full time span since the beginning of measurement in 2019 (Supp Mat S3 and Supp Mat S4). In summary, we show that the Contact Index CXof an unobserved complete network G can be retrieved from a network sample G^* obtained under the described sampling scheme according to

$$CX - 1 = \frac{CX^* - 1}{pq_{\text{eff}}},\tag{1}$$

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

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

235 However, network sampling destroys topological 198 236 information about underlying complete networks 199 237 (Figure 1(f)); the success of Horvitz-Thompson the-200 238 ory [21] to establish a connection between original 201 239 and sample networks relies in the use of weighted 202 links (Supp Mat S3). To establish the same con-203 241 nection for unweighted networks, we devised a 204 242 Bayesian approach which identifies missing topo-205 243 logical information as the weight distribution for 206 244 existing links in the complete network P(w|w > 0)207 245 and defines the edge sampling probability as 208

$$q_{\text{eff}} = P(w^* > 0 | w > 0) = 1 - G_{w|w>0}(1-q), (2)^{-247}$$

where $G_{w|w>0}(\xi) = \sum_{w>0} P(w|w>0)\xi^w$ is the ²⁴⁹ probability generating function of P(w|w>0) ²⁵⁰ (Supp Mat S3). We find that available com- ²⁵¹ plete real-world networks in various contexts [17, ²⁵² 18, 19] appear to show strikingly similar weight ²⁵³ distributions (Figure 1(g)), which suggests a universal shape of P(w|w > 0) also applicable to Here, "complete" refers to the our problem. 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 exponents [25, 26] (Figure 1(g)), a repeatedly demonstrated feature of complex networks [27] and beyond [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 confounded 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 beginning of crowd-sourcing in 2019, despite the timedependent 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 under normal conditions, as shown by the Fall and Christmas breaks in 2019, thus showing a reduction 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 altered 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 CXunderwent several periods of spiking (partial or

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

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

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

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

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

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

The effective reproduction number $R_{\rm eff}$ 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 physiological processes involved in transmission and, together, define the intrinsic transmission efficiency (per contact) $T = U \cdot \tau$.

 $\frac{\langle k^2 \rangle}{\langle k \rangle}$ Furthermore, as we assume CXreplaces $\langle k \rangle$, 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 $R_{\rm eff}$ data at the optimal time delay of $\Delta t = 16$ days (Figure S1(a, left inset) and Supp Mat S5). Upon interpreting $R_{WT}(CX) \equiv a + b \cdot CX$ as the wild-type specific reproduction number, we have that

$$R_{\rm eff} = R_{\rm WT}(CX) \cdot T, \tag{3}$$

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

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

378

381

382

384

385

T as random fluctuations from the measurement of $_{374}$ R_{eff} and capture actual trends by $\langle T \rangle$, centered av- $_{375}$ erages over sliding time windows of 2 months (Supp $_{376}$ Mat S5).

338 3.3. Epidemic evolution of relative SARS-CoV-2 379 339 transmissibility 380

The evolution of relative SARS-CoV-2 transmissibility $\langle T \rangle$ is shown in Figure 2(b). This time series reenacts the various phases of the SARS-CoV-2 pandemic:

Relative SARS-CoV-2 transmissibility $\langle T \rangle$ is ap-344 proximately equal to unity throughout 2020, an 345 387 initial period purely driven by unperturbed wild-346 388 type epidemics that we used to "calibrate" CX and 347 $R_{\rm eff}$ which evolve on shorter timescales. It sub-348 sequently follows a tug-of-war pattern shaped by 349 391 alternating epidemic forces beyond contacts: im-350 mune escape variants and development of popula-351 tion immunity through infection and vaccination. 352 394 Three waves of increased relative transmissibility 353 are explained by the takeover of fitter virus lin-354 396 eages (Figure 2(b)), specifically alpha (spring 2021), 355 307 delta (summer 2021) and omicron BA.1/BA.2 (win-356 ter 2021/22). We hypothesize that subsequent re-357 laxation of $\langle T \rangle$ after each wave may be attributed to 358 400 natural immunity, while the superposed long-term 359 downward trend may be explained by the additional 360 402 immunity acquisition through (initial and booster) 361 vaccination campaigns. Interestingly, the effect of 362 omicron BA.4/BA.5 takeover in summer 2022 on 403 363 $\langle T \rangle$ is nowhere close to those of previous variants. 364

Comparing correlations with different parame- 404 365 ters rules out the possibility that the measured $\langle T \rangle$ 405 366 is shaped by factors confounding the reproduction ⁴⁰⁶ 367 numbers or CX values (Figure S1(b,c) and Supp ⁴⁰⁷ 368 Mat S5). These possible confounders include viral ⁴⁰⁸ 369 prevalence, CX itself through higher-order effects 409 370 from network sampling not captured by our mod-410 371 eling and other topological network features (such 372 as clustering, small-world properties) as well as $R_{\rm eff}$ 373

itself through changes in testing strategies and systematic under-reporting of infections [35]. For instance, testing individuals indiscriminately versus focusing test capacities on suspected infection cases may lead to incomparable snapshots of ongoing infection dynamics. Overall, strong positive correlation is exclusively observed between $\langle T \rangle$ and variant dynamics (Figure S1(b,c)) [36]. In this analysis, we use test positivity [37] and results from local prevalence studies [38] as proxies for overall prevalence. Also, we neglect possible effects from network sampling on different topological measures [39, 40], but we expect trends to be conserved as long as the sampling process remains unchanged.

We note the absence of seasonal oscillations in $\langle T \rangle$ as well as clear signatures of mask mandates (in effect across many social contexts between April 2020 and April 2022). A seasonal oscillation in $\langle T \rangle$, larger values in winter and smaller values in summer, might be expected from the shift of human activity between in- and outdoor settings. Also, previous research established the effectiveness of mask usage at reducing transmission of respiratory diseases (reviewed in [41]). Overall, our results suggest that, at least in the epidemic stage of SARS-CoV-2, infection rates were predominantly driven by the strong variability in contacts as well as the repeated emergence of more transmissible variants, in line with previous findings [42, 43, 44].

3.4. Forecast of infection level and trend changes

The challenge of epidemic forecast consists in the accurate prediction of current and future reproduction numbers R_{eff} . Using the rationale that trends in infection levels carry the combined signature of trends in contact and relative transmissibility levels, we propose to construct predictions according to

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

459

461

462

463

466

467

470

471

476

477

470

480

483

485

486

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

Figure 3(a) shows R_{true} together with data from 418 infection surveillance, both plotted with respect to 419 their date of recording (assuming real-time CX420 This shows how our prediction measurement). 421 overall anticipates current epidemic trends that 422 464 are observed via infection surveillance only about 423 $\Delta t = 2 - 3$ weeks later. Thus, we propose to use 424 our method as a tool for real-time infection surveil-425 lance. 426

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

Most importantly, trend changes in epidemic 487 447 driving forces such as $\langle T \rangle$ and CX are indicators of 488 448 new phases in an epidemic. Timely detection of new 489 449 trends in these time series, e.g. using anomaly de-490 450 tection methods, can provide valuable information 491 451

to estimate the risk of upcoming epidemic waves and to predict their nature – whether dynamics is fueled by contacts or increased transmission efficiency. Such trend detection is potentially easier to achieve but equally informative than the ability to accurately predict infection surveillance. The onset of rising trends could shape decision-making with regard to the effectiveness of health policies, pharmaceutical and non-pharmaceutical ine.g. 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 Influenza, respectively, akin to trends in stock prices. For SARS-CoV-2, trend changes are timely indicators 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 (Figure S2(b)).

Discussion

We presented a simple, yet insightful quantitative method for a data-driven decomposition of overall epidemic dynamics into contact-related and transmission efficiency-related contributions. It relies on both the availability of infection surveillance data as well as crowd-sourced GPS location data to detect and quantify physical proximity between susceptible individuals. Its appeal resides in the merely bivariate yet highly informative projection of epidemics paving the way towards timely identification of driving forces in an ongoing epidemic - human versus viral factors - and possibly effective mitigation strategies – pharmaceutical versus non-pharmaceutical.

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

or reproduction numbers, thus taking our previ- 533 492 ously described short-term forecast further [5]. Yet, 534 493 a timely detection of trend changes could reliably 535 494 forecast upcoming waves and their nature without 536 495 the necessity to accurately predict infection surveil- 537 496 lance data. These tools can lead towards a more 538 497 strategic approach to epidemic mitigation and po- 539 498 tentially save lives by reducing the spread of deadly 540 499 diseases. 500 541

Results from the presumably most systematically 542 501 tracked epidemic to date, SARS-CoV-2, draw the 502 544 picture of co-evolution within the virus-host rela-503 545 tion: Increasing immunity levels in the host pop-504 ulation alternate with step-wise adaptation of the 546 505 virus through immune-escape variants. Other fre-547 506 quently discussed factors, including mask policies 548 507 and seasonality, are presumably still below the cur-508 rent statistical resolution of our method, defined by 509 the sampling noise in the CX and R_{eff} time series. 550 510 Moreover, a larger impact of seasonal variation is 511 expected in the endemic phase of SARS-CoV-2 [46]. 551 512

Our method is broadly applicable to airborne 513 contagions beyond SARS-CoV-2, but depends on 514 the availability of infection surveillance and crowd-515 sourcing strategies that remain persistent over ex-516 tended amounts of time. Changes in testing strat-517 557 egy can lead to signal and biases unrelated to un-518 derlying epidemic driving forces [35]. More cru-519 558 cially, systematic infection surveillance is not im-520 plemented beyond the case of SARS-CoV-2. We 521 550 illustrated a framework to correct for the effect 560 522 561 of varying sampling depth in the contact network. 523 Yet, higher-order effects in the signal can occur as 524 563 a result of sampling aspects not captured by our 525 56/ mathematical modeling. In order to ensure valid 526 566 prognoses through our method, we advocate for sys-527 tematic and persistent crowd-sourcing and infection 528 surveillance strategies across a variety of diseases 569 529 with epidemic potential. 570 530

Geographical resolution of our forecast method 531 572 is currently limited by the sampling depth, as the 532 573 estimation especially of higher moments of degree distributions P(k) becomes increasingly difficult as smaller portions of the network are available. A higher spatial resolution of contact and relative transmissibility levels, with potential to locate the origin of new variants of concern and define locally targeted mitigation strategies, can be achieved by e.g. increasing the panel of app users.

Our analysis assumes statics, but actual contact networks are dynamic in nature [47, 48]: While some contacts are frequently repeated (e.g. between household members), other contacts are randomly redrawn on each occasion (e.g. in public transportation), with implications for epidemic spread [49, 50]. Our method can be improved by analyzing contact data in light of existing models of dynamic networks [51, 48].

Acknowledgment

This work was supported by grants from the Federal Government of Germany through the Federal Ministry for Economic Affairs and Climate Action (BMWK) for the project DAKI-FWS (01MK21009A) and the Federal Ministry of Education and Research (BMBF) for the project Optim-Agent (031L0299).

References

552

554

555

562

- [1] A. Rodríguez, H. Kamarthi, P. Agarwal, J. Ho, M. Patel, S. Sapre, B. A. Prakash, Data-centric epidemic forecasting: A survey, arXiv preprint arXiv:2207.09370 (2022).
- [2] A. Maxmen, Has covid taught us anything about pandemic preparedness?, Nature 596 (2021) 332-335.
- [3] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, Vespignani, Epidemic processes in complex Α. Rev. Mod. Phys. 87 (2015) 925 - 979.networks. doi:10.1103/RevModPhys.87.925. URL https://link.aps.org/doi/10.1103/ RevModPhys.87.925
- [4]T. Alamo, D. G. Reina, P. Millán Gata, V. M. Preciado, G. Giordano, Data-driven methods for present and future pandemics: Monitoring, modelling and managing,

628

629

- 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/
```

```
577 article/pii/S1367578821000419
```

- 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
- 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
- arXiv:https://www.pnas.org/doi/pdf/10.1073
 pnas.2026731118, doi:10.1073/pnas.2026731118.
- 584
 pnas.2026731118, doi:10.1073/pnas.2026731118.
 636

 585
 URL https://www.pnas.org/doi/abs/10.1073/pnas.
 637

 586
 2026731118
 638
- [6] M. E. J. Newman, J. Park, Why social networks are 639
 different from other types of networks, Phys. Rev. E 68 640
 (2003) 036122. doi:10.1103/PhysRevE.68.036122. 641
- 590 URL https://link.aps.org/doi/10.1103/PhysRevE. 642
 591 68.036122 643
- 592[7] M. Keeling, The implications of network structure644593for epidemic dynamics, Theoretical population biology64559467 (1) (2005) 1–8.646
- [8] C. Moore, M. E. J. Newman, Epidemics and percolation 647
 in small-world networks, Phys. Rev. E 61 (2000) 5678- 648
 5682. doi:10.1103/PhysRevE.61.5678. 649
- 598
 URL https://link.aps.org/doi/10.1103/PhysRevE.
 650

 599
 61.5678
 651
- [9] M. Barthélemy, A. Barrat, R. Pastor-Satorras, 652 600 A. Vespignani, Dynamical patterns of epidemic out-653 601 breaks in complex heterogeneous networks, Jour- 654 602 nal of Theoretical Biology 235 (2) (2005) 275-288. 655 603 604 doi:https://doi.org/10.1016/j.jtbi.2005.01.011. 656 URL https://www.sciencedirect.com/science/ 657 605 article/pii/S0022519305000251 606 658
- [10] Y. Moreno, R. Pastor-Satorras, A. Vespignani, Epi- 659
 demic outbreaks in complex heterogeneous networks, 660
 The European Physical Journal B-Condensed Matter 661
 and Complex Systems 26 (4) (2002) 521–529. 662
- [11] R. M. May, R. M. Anderson, M. E. Irwin, R. M. 663
 Anderson, J. M. Thresh, The transmission dynamics 664
 of human immunodeficiency virus (hiv), Philosophical 665
 Transactions of the Royal Society of London. B, Bi- 666
 ological Sciences 321 (1207) (1988) 565-607. arXiv: 667
 https://royalsocietypublishing.org/doi/pdf/10. 668
- 617 1098/rstb.1988.0108, doi:10.1098/rstb.1988.0108. 669
- 618
 URL https://royalsocietypublishing.org/doi/abs/
 670

 619
 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
- [13] S. L. Feld, Why your friends have more friends than 674
 you do, American Journal of Sociology 96 (6) (1991) 675
 1464–1477. arXiv:https://doi.org/10.1086/229693, 676
 doi:10.1086/229693. 677

URL https://doi.org/10.1086/229693

- [14] J. M. Read, K. T. Eames, W. J. Edmunds, Dynamic social networks and the implications for the spread of infectious disease, Journal of The Royal Society Interface 5 (26) (2008) 1001–1007.
- [15] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, W. J. Edmunds, Social contacts and mixing patterns relevant to the spread of infectious diseases, PLOS Medicine 5 (3) (2008) 1–1. doi:10.1371/journal.pmed.0050074.

URL https://doi.org/10.1371/journal.pmed. 0050074

- [16] G. E. Leventhal, R. Kouyos, T. Stadler, V. von Wyl, S. Yerly, J. Böni, C. Cellerai, T. Klimkait, H. F. Günthard, S. Bonhoeffer, Inferring epidemic contact structure from phylogenetic trees, PLOS Computational Biology 8 (3) (2012) 1-10. doi:10.1371/journal.pcbi.1002413. URL https://doi.org/10.1371/journal.pcbi. 1002413
- [17] R. Pung, J. A. Firth, L. G. Spurgin, V. J. Lee, A. J. Kucharski, Using high-resolution contact networks to evaluate sars-cov-2 transmission and control in largescale multi-day events, Nature communications 13 (1) (2022) 1–11.
- [18] P. Sapiezynski, A. Stopczynski, D. D. Lassen, S. Lehmann, Interaction data from the copenhagen networks study, Scientific Data 6 (1) (2019) 315.
- [19] S. M. Kissler, P. Klepac, M. Tang, A. J. Conlan, J. R. Gog, Sparking "the bbc four pandemic": Leveraging citizen science and mobile phones to model the spread of disease, bioRxiv (2020). arXiv:https://www.biorxiv. org/content/early/2020/05/12/479154.full.pdf, doi:10.1101/479154.

URL https://www.biorxiv.org/content/early/2020/ 05/12/479154

[20] F. W. Crawford, S. A. Jones, M. Cartter, S. G. Dean, J. L. Warren, Z. R. Li, J. Barbieri, J. Campbell, P. Kenney, T. Valleau, O. Morozova, Impact of close interpersonal contact on covid-19 incidence: Evidence from 1 year of mobile device data, Science Advances 8 (1) (2022) eabi5499. arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.abi5499, doi:10.1126/sciadv.abi5499.

URL https://www.science.org/doi/abs/10.1126/ sciadv.abi5499

- [21] E. D. Kolaczyk, Statistical Analysis of Network Data, Springer New York, NY, 2009. doi:https://doi.org/ 10.1007/978-0-387-88146-1.
- [22] P. Hu, W. C. Lau, A survey and taxonomy of graph

678		sampling, CoRR abs/1308.5865 (2013). arXiv:1308.	730	
679		5865.	731	
680		URL http://arxiv.org/abs/1308.5865	732	
681	[23]	M. A. Serrano, M. Boguñá, R. Pastor-Satorras, Corre-	733	[:
682		lations 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	[94]	74.055101 A Barrat M Barthálanas B Baston Satarrag	737	
686	[24]	A. Barrat, M. Barthelemy, R. Pastor-Satorras,	738	
687		A. vespignani, The architecture of complex weighted	739	
680		Sciences 101 (11) (2004) 3747-3752 arXiv:https:	740	
600		//www.pnas.org/doi/pdf/10_1073/pnas_0400087101	741	[5
691		doi:10.1073/pnas.0400087101.	743	Į¢
692		URL https://www.pnas.org/doi/abs/10.1073/pnas.	744	
693		0400087101	745	[3
694	[25]	H. Ebel, LI. Mielsch, S. Bornholdt, Scale-free topology	746	Ľ
695		of e-mail networks, Phys. Rev. E 66 (2002) 035103. doi:	747	[3
696		10.1103/PhysRevE.66.035103.	748	-
697		URL https://link.aps.org/doi/10.1103/PhysRevE.	749	
698		66.035103	750	
699	[26]	M. E. J. Newman, Scientific collaboration networks. i.	751	
700		network construction and fundamental results, $\ensuremath{\operatorname{Phys.}}$	752	[3
701		Rev. E 64 (2001) 016131. doi:10.1103/PhysRevE.64.	753	
702		016131.	754	
703		URL https://link.aps.org/doi/10.1103/PhysRevE.	755	
704		64.016131	756	_
705	[27]	R. Guimera, S. Mossa, A. Turtschi, L. N. Amaral,	757	[:
706		The worldwide air transportation network: Anoma-	758	
707		lous centrality, community structure, and cities' global	759	
708		roles, Proceedings of the National Academy of Sciences	760	
709	[90]	102 (22) (2005) 7794–7799.	761	
710	[28]	A. Clauset, C. R. Shalizi, M. E. J. Newman, Power-law	762	[f
711		(2000) 661 702 arVirubttra: (/doi org/10.1127/	763	Į
712		070710111 doi:10.1137/070710111	764	
713		UBL https://doi_org/10_1137/070710111	765	
715	[29]	J-D. J. Han, D. Dupuy, N. Bertin, M. E. Cusick.	767	
716	[=0]	M. Vidal. Effect of sampling on topology predictions	768	[4
717		of protein-protein interaction networks, Nature biotech-	769	Ľ
718		nology 23 (7) (2005) 839–844.	770	
719	[30]	R. Perline, Strong, weak and false inverse power laws,	771	
720		Statistical Science 20 (1) (2005) 68–88.	772	
721		$\mathrm{URL}\ \mathtt{http://www.jstor.org/stable/20061161}$	773	[4
722	[31]	T. Hale, N. Angrist, R. Goldszmidt, B. Kira, A. Pether-	774	
723		ick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hal-	775	
724		las, S. Majumdar, et al., A global panel database of pan-	776	
725		demic policies (oxford covid-19 government response	777	
726		tracker), Nature human behaviour 5 (4) (2021) 529–	778	
727	F = - 7	538.	779	
728	[32]	C. Betsch, S. Eitze, P. Sprengholz, L. Korn, P. Sham-	780	
729		srizi, M. Geiger, E. Sievert, L. Lehrer, M. Jenny, Zusam-	781	4
			10	

menfassung und empfehlungen welle 69 (2022).

URL https://projekte.uni-erfurt.de/cosmo2020/ web/summary/69/

- [33] C. Betsch, L. Wieler, M. Bosnjak, M. Ramharter, V. Stollorz, S. Omer, L. Korn, P. Sprengholz, L. Felgendreff, S. Eitze, P. Schmid, Germany covid-19 snapshot monitoring (cosmo germany): Monitoring knowledge, risk perceptions, preventive behaviours, and public trust in the current coronavirus outbreak in germany (Mar. 2020). doi:10.23668/psycharchives.2776. URL https://psycharchives.org/index.php/en/ item/e5acdc65-77e9-4fd4-9cd2-bf6aa2dd5eba
- [34] M. an der Heiden, Sars-cov-2-nowcasting und -r-schaetzung (Jan. 2023). doi:10.5281/zenodo.7571376. URL https://doi.org/10.5281/zenodo.7571376
- [35] H. Rossman, E. Segal, Nowcasting the spread of sarscov-2, Nature microbiology 7 (1) (2022) 16–17.
- [36] R. Koch-Institut, Anzahl und anteile von voc und voi in deutschland (2023). URL https://www.rki.de/DE/Content/InfAZ/N/ Neuartiges_Coronavirus/Daten/VOC_VOI_Tabelle. xlsx?_blob=publicationFile
- [37] E. Mathieu, H. Ritchie, L. Rodés-Guirao, C. Appel, C. Giattino, J. Hasell, B. Macdonald, S. Dattani, D. Beltekian, E. Ortiz-Ospina, M. Roser, Coronavirus pandemic (covid-19) (2020).

 $\mathrm{URL}\ \mathtt{https://ourworldindata.org/coronavirus}$

[38] H. Neuhauser, N. Buttmann-Schweiger, J. Fiebig, C. Poethko-Müller, F. Prütz, G. Sarganas Margolis, R. Thamm, M. Zimmermann, Observatorium serologischer Studien zu SARS-CoV-2 in Deutschland (Sep. 2022). doi:10.5281/zenodo.7043025.

 $\mathrm{URL}\ \mathtt{https://doi.org/10.5281/zenodo.7043025}$

- [39] J. D. Noh, Percolation transition in networks with degree-degree correlation, Phys. Rev. E 76 (2007) 026116. doi:10.1103/PhysRevE.76.026116. URL https://link.aps.org/doi/10.1103/PhysRevE. 76.026116
- [40] S. H. Lee, P.-J. Kim, H. Jeong, Statistical properties of sampled networks, Phys. Rev. E 73 (2006) 016102. doi:10.1103/PhysRevE.73.016102. URL https://link.aps.org/doi/10.1103/PhysRevE. 73.016102
- [41] J. T. Brooks, J. C. Butler, Effectiveness of Mask Wearing to Control Community Spread of SARS-CoV-2, JAMA 325 (10) (2021) 998-999. arXiv:https://jamanetwork.com/journals/ jama / articlepdf / 2776536 / jama \ _brooks \ _2021 \ _it \ _210006 \ _1631033869 . 97869 . pdf, doi:10.1001/jama.2021.1505.

URL https://doi.org/10.1001/jama.2021.1505

[42] F. Balloux, C. Tan, L. Swadling, D. Richard, C. Jenner,

839

5.021005

- 782 M. Maini, L. van Dorp, The past, current and future 834
- repidemiological dynamic of SARS-CoV-2, Oxford Open 835
- 784 Immunology 3 (1) (06 2022). arXiv:https://academic. 836
- 785 oup.com/ooim/article-pdf/3/1/iqac003/48744431/ 837 iqac003.pdf, doi:10.1093/oxfimm/iqac003. 838
- 786 iqac003.pdf, doi:10.1093/oxfimm/iqac003.
- 787 URL https://doi.org/10.1093/oxfimm/iqac003
- G. P. Guy Jr, F. C. Lee, G. Sunshine, R. McCord, 840 788 [43]789 M. Howard-Williams, L. Kompaniyets, C. Dunphy, 841 M. Gakh, R. Weber, E. Sauber-Schatz, et al., Asso-842 790 ciation of state-issued mask mandates and allowing on- 843 791 premises restaurant dining with county-level covid-19 844 792 case and death growth rates—united states, march 1-845 793 december 31, 2020, Morbidity and Mortality Weekly 846 794 Report 70 (10) (2021) 350. 847 795
- [44] D. K. Chu, E. A. Akl, S. Duda, K. Solo, S. Yaacoub, 848 796 H. J. Schünemann, D. K. Chu, E. A. Akl, A. El-harakeh, 849 797 A. Bognanni, T. Lotfi, M. Loeb, A. Hajizadeh, A. Bak, 850 79 A. Izcovich, C. A. Cuello-Garcia, C. Chen, D. J. Har-851 799 ris, E. Borowiack, F. Chamseddine, F. Schünemann, 852 800 G. P. Morgano, G. E. U. Muti Schünemann, G. Chen, 853 801 H. Zhao, I. Neumann, J. Chan, J. Khabsa, L. Hneiny, 854 802 L. Harrison, M. Smith, N. Rizk, P. Giorgi Rossi, P. Abi- 855 803
- Hanna, R. El-khoury, R. Stalteri, T. Baldeh, T. Piggott, 856
- Y. Zhang, Z. Saad, A. Khamis, M. Reinap, S. Duda, 857
 K. Solo, S. Yaacoub, H. J. Schünemann, Physical dis- 858
- tancing, face masks, and eye protection to prevent 859
- person-to-person transmission of sars-cov-2 and covid-
- ⁸⁰⁹ 19: a systematic review and meta-analysis, The Lancet
- 810 395 (10242) (2020) 1973-1987. doi:https://doi.org/
- ⁸¹¹ 10.1016/S0140-6736(20)31142-9.
- 812 URL https://www.sciencedirect.com/science/ 813 article/pii/S0140673620311429
- 814 [45] R. Koch-Institut, Survstat@rki 2.0 (2023).
- 815 URL https://survstat.rki.de/
- J. P. Townsend, A. D. Lamb, H. B. Hassler, [46]816 Sah, A. A. Nishio, C. Nguyen, A. D. Tew, Р. 817 A. P. Galvani, A. Dornburg, Projecting the sea-818 sonality of endemic covid-19, medRxiv (2022). 819 arXiv: https://www.medrxiv.org/content/early/ 820 2022/10/07/2022.01.26.22269905.full.pdf, doi:10. 821 1101/2022.01.26.22269905. 822
- URL https://www.medrxiv.org/content/early/2022/
 10/07/2022.01.26.22269905
- [47] V. Sekara, A. Stopczynski, S. Lehmann, Fundamental structures of dynamic social networks,
 Proceedings of the National Academy of Sciences
 113 (36) (2016) 9977–9982. arXiv:https://
 www.pnas.org/doi/pdf/10.1073/pnas.1602803113,
 doi:10.1073/pnas.1602803113.
- URL https://www.pnas.org/doi/abs/10.1073/pnas.
 1602803113
- 833 [48] P. Holme, J. Saramäki, Temporal networks, Physics Re-

ports 519 (3) (2012) 97-125, temporal Networks. doi: https://doi.org/10.1016/j.physrep.2012.03.001. URL https://www.sciencedirect.com/science/ article/pii/S0370157312000841

- [49] J. Enright, R. R. Kao, Epidemics on dynamic networks, Epidemics 24 (2018) 88-97. doi:https://doi.org/10. 1016/j.epidem.2018.04.003. URL https://www.sciencedirect.com/science/ article/pii/S1755436518300173
- [50] E. Valdano, L. Ferreri, C. Poletto, V. Colizza, Analytical computation of the epidemic threshold on temporal networks, Phys. Rev. X 5 (2015) 021005. doi:10.1103/PhysRevX.5.021005. URL https://link.aps.org/doi/10.1103/PhysRevX.
- [51] X. Zhang, C. Moore, M. E. Newman, Random graph models for dynamic networks, The European Physical Journal B 90 (10) (2017) 1–14.
- [52] N. Pitsianis, D. Floros, A.-S. Iliopoulos, X. Sun, Sg-tsne-π: Swift neighbor embedding of sparse stochastic graphs, Journal of Open Source Software 4 (39) (2019) 1577.
- [53] N. Pitsianis, A.-S. Iliopoulos, D. Floros, X. Sun, Spaceland embedding of sparse stochastic graphs, in: 2019 IEEE High Performance Extreme Computing Conference (HPEC), 2019, pp. 1–8. doi:10.1109/HPEC.2019. 8916505.



Figure 1: Contact networks: definition, measurement and inference. (a) Co-location of 2 (or more) devices i and jwithin 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 $w_{ij} \in \{0, 1, 2, ...\}$ 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 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).