HUNGARY IN MASKS/"MASZK" IN HUNGARY

Márton Karsai,[1,2] *Júlia Koltai,*[3,4,1] *Orsolya Vásárhelyi,*[5,1] *Gergely Röst*[6*]

INTRODUCTION

Social interactions represent one of the most important routes of transmission of COVID-19 as they influence the potential patterns of diffusion of infection throughout different segments of the population. Despite their utmost importance, the scientific community is currently lacking data collection methods that record social interactions dynamically and in detail, and in a privacy-respecting, representative way, even on an aggregated level. Here we summarize the motivation, methodology, and some early results of a coordinated process of data collection in Hungary designed to track the social mixing patterns of people in different age groups in real time during the pandemic. The Hungarian Data Provider Questionnaire (MASZK[7]) was released in late March 2020 during the initial phase of the COVID-19 outbreak in Hungary. This is an ongoing effort to anonymously collect age contact matrices of a voluntary population online. Moreover, it is accompanied with a nationally representative data collection campaign via telephone survey to ensure data quality. This unique process

1 Central European University, Department of Network and Data Science, Vienna, A-1100, Austria.

2 Alfréd Rényi Institute of Mathematics, 1053 Budapest, Hungary.

3 Centre for Social Sciences, Hungarian Academy of Sciences Centre of Excellence, Computational Social Science – Research Center for Educational and Network Studies, Budapest, H-1097, Hungary.

4 Faculty of Social Sciences, Eötvös Loránd University, Budapest, H-1117, Hungary.

5 University of Warwick, Center for Interdisciplinary Methodologies, Coventry, CV4 7AL, United Kingdom.

6 Bolyai Institute, University of Szeged, Szeged, H-6720, Hungary. Corresponding author: rost@math.u-szeged

7 Acronym for Magyar Adatszolgáltató Kérdőív.

of combined data collection is an important step towards developing more precise modelling of the spread of the epidemic, and could thus contribute to the reduction of the medical and economic burden of the pandemic in Hungary.

As the first wave of the global COVID-19 pandemic reached Hungary in March 2020, early operative action was called for to analyze the epidemiological situation of Hungary and project potential outcome scenarios for decision makers. A working group called the COVID-19 Mathematical Modelling and Epidemiological Analysis Task Force was thus established. This multi-disciplinary group of experts involves researchers from most major universities and research institutions in the country who are working in various fields, including epidemiology, public health, medicine, health management, mathematics, network science, biostatistics, computer science, data science, and sociology. The task force has continuously worked together with the Ministry of Innovation and Technology during the pandemic, creating reports, analyses, and strategic documents, with the goal of providing expert advice to decision makers based on the best available scientific evidence. This joint effort was accompanied by an unprecedented wave of data sharing, whereby health authorities, major telecommunication companies, the Hungarian Central Statistical Office, and other data shareholders such as global online social networks (e.g., Facebook D4G,[8] and Google Community Mobility Report[9]) opened up their resources to help the fight the pandemic with data. Although these datasets provided precise descriptions about the demographic factors and the actual mobility of the population separately, they all fell short in terms of capturing the spatially, temporally, and demographically detailed age-mixing patterns of the country's population. In epidemiological models, such information is conventionally summarized using aggregated age contact matrices that code the average number of interactions between people of different age groups. Similar matrices were earlier estimated for Hungary (Prem et al. 2017), and have also been collected using contact diaries (Dávid et al. 2016), but related changes during the actual pandemic due to national lock-down and social distancing measures were entirely unknown. The Hungarian Data Provider Questionnaire (MASZK) was developed within the frame of the COVID-19 Mathematical Modelling and Epidemiological Analysis Task Force by scientists and software engineers,[10] specifically with the purpose of continuously monitoring changes in social mixing patterns. The survey anonymously collects demographic

8 Facebook – data for development – COVID-19: https://dataforgood.fb.com/docs/covid19/ (date of access 2020.09.28).

9 Google – community mobility report: https://www.google.com/covid19/mobility/ (date of access 2020.09.28).

10 https://covid.sed.hu/tabs/staff (date of access 2020.09.28)

information about respondents, as well as data about their domicile, education, health-, and family structure. Its primary goal is to record daily changes in age contact matrices. Furthermore, it also helps to monitor the level of compliance with recommended self-protection measures, and the social, physical, and mental well-being of individuals. Although the data collection is ongoing, as of now (October 2020), it has involved more than two percent of the population of Hungary, with the collection of over 320,000 questionnaires from more than 210,000 unique participants.

However, due to voluntary participation, the online sample has not provided a representative description of the Hungarian population: Young and middle-aged female adults with university degrees living in larger cities are over-represented compared to their number in the national census from 2011.[11] To overcome these biases, starting in April 2020 the same questionnaire has been conducted on a nationally representative sample of 1,500 people every month via telephone survey (CATI) by a market research company. Moreover, to validate the results of the traditional surveys, contact diaries (Dávid et al. 2016) were collected on a nationally representative sample of 1,000 people. Through a combined analysis of the representative and non-representative datasets, and also census data, we identified the most severe biases characterizing the online survey data and built a pipeline (Koltai et al. 2021) using iterative proportional fitting (Bishop et al. 2007) to weight the non-representative online matrices. As a result, we obtained close-to-representative contact matrices from the non-representative online data, while maintaining the advantages of the online data collection, such as scalability of size and observation time, and cost efficiency. In what follows, after a short summary of the questionnaire we summarize some first findings about the aforementioned contact matrices and provide a short discussion that elaborates on the significance and future directions of our study.

RESULTS

The MASZK questionnaire

The goal of the MASZK questionnaire was not the one-time screening of social mixing, but to record real-time, dynamic changes in contact patterns during the pandemic compared to the pre-COVID-19 reference period. To

11 http://www.ksh.hu/nepszamlalas/ (date of access 2020.09.28)

enhance engagement and minimize the turnover of respondents, the survey was designed in a modular way, so participants could provide data every day without being burdened. More specifically, we divided the survey into two parts: a static part of the questionnaire that was asked once – on the first occasion the respondent filled out the survey – that collected demographic information (age, gender, occupation, education), and identified the health profile (chronic/acute health conditions), and the household structure of the respondent. Furthermore, we asked about respondents' employment conditions, mobility, and commuting patterns, self-protection measures, and participants' attitude towards potential vaccination. In this part, we also recorded the age contact matrices of the participants concerning their typical weekdays and weekend during pre-COVID-19 times (before 13 March 2020). Such matrices were recorded by asking respondents to estimate the number of their social interactions with others of different age groups. Social interactions were defined as *proxy* contacts between people who had been closer to each other than two meters for longer than 15 minutes. Contact matrices for children (those aged under 18) were also collected, but only through the reporting of their parents, especially focusing on estimating the former's number of contacts at school and elsewhere.

The dynamic part of the questionnaire collected data about the participant's previous day. Respondents were asked to fill out this part of the questionnaire potentially every day. In this part, we asked about their contacts in different settings, such as outside of home, at work, or in the case of having visitors. Here we measured the number of contacts in two ways. In the first way, we used the above-described *proxy* definition, while we also asked about their *physical* contacts – i.e. if respondents had had skin-to-skin interaction without any protection (e.g. rubber gloves).[12] The same questions were asked about the under-aged children of the respondents.

Changes in contact patterns

The age contact matrices recorded on representative samples for the reference period and for two others during the pandemic are shown in Fig. 1a-c. Here, the reference period matrix (see Fig. 1a) revealed the expected structure, with a pronounced diagonal component indicating strong age-homophily effects,

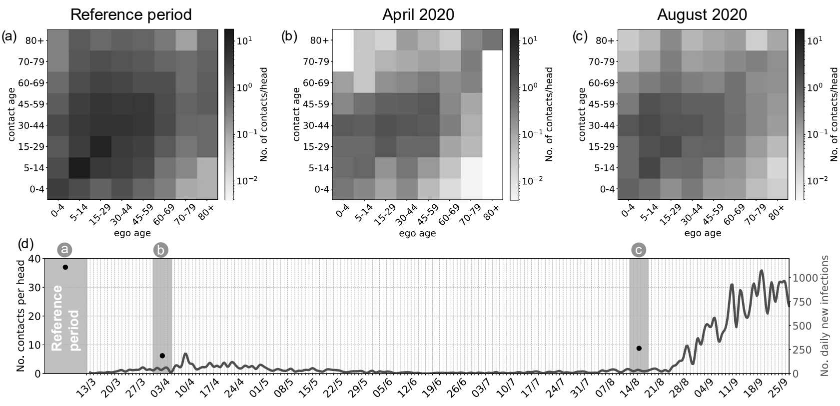
12 Surveillance definitions for COVID-19, European Centre for Disease Prevention and Control, https://www.ecdc.europa.eu/en/covid-19/surveillance/surveillance-definitions, (date of access 2020.09.28)

especially for children and young teenagers. However, young and middle-aged people appeared to be best connected to each other due to the age composition of families, and frequent contact between children, parents, and grandparents. Nevertheless, the oldest generation (80+) appeared to be the least well connected to very young children, as already observed in earlier studies (Mossong et al. 2008). The average number of social proxy connections per person was measured to be 36.985, as indicated by the black point in Fig. 1d during the reference period.

Following the emergence of the COVID-19 pandemic in Hungary, a two-month national lock-down was introduced to slow down and reduce the first wave of infections. This measure considerably reduced the number of social contacts of people, which are thought to be the main route of transmission of the disease. The compliance of individuals with these regulations is clearly reflected by the contact matrices recorded immediately after the start of the lock-down at the beginning of April 2020. As shown in Fig. 1b, the number of contacts radically decreased in every age group. This was moreover true for elderly people, who appeared to be the most vulnerable to the disease. Their number of contacts, especially with young children, declined by over an order of magnitude. Nevertheless, they still maintained contact with middle-aged people and others from their own age group. For society as a whole, the average number of contacts during this period decreased almost six-fold to 6.195, as indicated by the black point in Fig. 1d, corresponding to the April data collection period.

As in several other countries in Europe, a second COVID-19 infection wave occurred in Hungary at the beginning of September 2020, as was expected after the progressive relaxation of control measures (Röst et al. 2020). While the underlying reasons for the timing and the intensity of the second wave are not entirely understood, an increase in the number of social connections – which can be clearly seen from our data (see Fig. 1c) – may have contributed to this phenomenon. By the time of the observation in the middle of August 2020, the average number of social contacts had increased to 8.795 (see the black point in Fig. 1d in August). Moreover, mixing between different age groups had homogenized. The isolation of elderly people had declined and homophily was less observable due to the increase in mixing of the population. One reason for these more evenly mixed patterns may be the overlap with the vacation period – when people took holidays further from their homes, sometimes even abroad. These enhanced spatial mobility patterns, decay of social awareness, and increase in the number of social contacts may have induced or amplified the re-emergence of the pandemic in the Hungarian population, shown as the solid red line in Fig. 1d.

Figure 1: *Age contact matrices based on representative telephone survey data for (a) the reference period (before 13 March 2020); for (b) the early period of the COVID-19 pandemic; and (c) immediately before the second wave of the pandemic in Hungary. The X-axes of matrices indicate the age group of the participants, while Y-axes assign the age group of their proxy social contacts (for definition, see text). Color codes show (logarithmically) the average number of connections between age groups, normalized identically between the different matrices. Panel (d) shows the evolution of the number of daily infections in Hungary (red solid line) and the average number of contacts (black dots) during the observed periods (blue shaded rectangles).*



DISCUSSION

While Hungary is still facing the pandemic, just like almost every country on Earth, unprecedented efforts have been mounted to create a coordinated response to this crisis worldwide. The first wave of the COVID-19 outbreak was successfully suppressed in Hungary due to timely and strict interventions that resulted in a drastic reduction in the number of contacts (Röst et al. 2020). The resurgence of the virus in early fall made it evident that this pandemic will not only have medical consequences for humanity, but will also induce a long-lasting social and economic crisis (Fernandes 2020). There are several areas still associated with challenges and strong time pressure in relation to avoiding even more severe consequences thereof. These challenges include not only the development of an effective vaccination program, and the mitigation of the economic burden, but also precise and innovative solutions for modelling the spread of the disease that will generate insights into the processes governing the dynamics of COVID-19. In relation to the dissemination of the disease, the scientific community is facing new and unknown scenarios on a daily basis, impacted by spatial, demographic, and even cultural effects. Our contribution may be important in the sense that it provides a methodology for dynamically observing social mixing patterns, stratified by age. Our results provide key inputs for epidemic models, thereby helping build forecasts on better assumptions and up-to-date information about the social mixing underlying the transmission of the disease.

The *limitations* of our voluntary online data collection process were clear from the very beginning. In response, we developed a methodology (Koltai et al. 2021) which may be useful for scaling non-representative age contact matrices to increase their approximations of reality, and thus to more accurately describe the underlying population. This methodology involves identifying those factors that affect the contact patterns of people on representative national samples, and weighting the online matrices using census data related to these factors. These preliminary results may lead to the adoption of similar data collection methods by the scientific community, increasing the accuracy of the modelling of the pandemic.

After completion of this project, data will be made available for further research purposes upon request.

Acknowledgements

The authors are very thankful to the MASZK COVID-19 development team led by Vilmos Bilicki from the Department of Software Development at the University of Szeged. This work was carried out within the framework of the Hungarian National Development, Research, and Innovation (NKFIH) Fund 2020-2.1.1-ED-2020-00003. The work of Júlia Koltai was funded by the Premium Postdoctoral Grant of the Hungarian Academy of Sciences. Gergely Röst was also supported by NKFIH FK 124016.

REFERENCES

Bishop, Yvonne M., Fienberg, Stephen E., and Holland, Paul W. (2007) *Discrete multivariate analysis: theory and practice*, New York, Springer-Verlag

Dávid, Beáta, Huszti, Éva, Barna, Ildikó, and Fu, Yang-chih (2016) Egocentric contact networks in comparison: Taiwan and Hungary. *Social Networks*, Vol. 44, pp. 253–265, https://doi.org/10.1016/j.socnet.2015.10.001

Fernandes, Nuno (2020) Economic Effects of Coronavirus Outbreak (COVID-19) on the World Economy, *IESE Business School Working Paper*, No. WP-1240-E, http://dx.doi.org/10.2139/ssrn.3557504

Koltai, Júlia, Vásárhelyi, Orsolya, Röst, Gergely, and Karsai, Márton (2021) Monitoring behavioural responses during pandemic via reconstructed contact matrices from online and representative surveys. *arXiv:2102.09021* (submitted).

Mossong Joel, Hens, Niel, Jit, Mark, Beutels, Philippe, Auranen, Kari, et al. (2008) Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Medicine* 5(3): e74. https://doi.org/10.1371/journal.pmed.0050074

Prem, Kiesha, Cook, Alex R., and Jit, Mark (2017) Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Computational Biology*, Vol. 13 No.9:e1005697, https://doi.org/10.1371/journal.pcbi.1005697