grippeNET App: Enhancing Participatory Influenza Monitoring Through Mobile Phone Sensors

Marco Hirsch
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany
marco.hirsch@dfki.de

Antoine Flahault
Institute of Global Health, Geneva
University of Geneva, Switzerland
antoine.flahault@unige.ch

Olivia Woolley-Meza
ETH Zurich, Switzerland
olivia.woolley@gess.ethz.ch

Paul Lukowicz
German Research Center for Artificial Intelligence (DFKI)
University of Kaiserslautern, Germany
paul.lukowicz@dfki.de

Daniela Paolotti
ISI Foundation, Torino, Italy
daniela.paolotti@isi.it

UbiComp/ISWC'18 Adjunct,, October 8–12, 2018, Singapore, Singapore
ACM 978-1-4503-5966-5/18/10.
https://doi.org/10.1145/3267305.3274171

Abstract
We describe the vision, implementation and initial deployment experience with an App based crowd sourcing system for the analysis and prediction of the influence of aggregated individual behavior patterns on the spread of infectious diseases. The App builds on InfluenzaNet [19], a well established European participatory network of self reporting web platforms for monitoring Influeza-Like-Ilnesses (ILI) and goes starting from the current practice of explicit self reporting towards the use of sensor derived behavior information. The App has been cleared with the Regional Research Ethics Committee (CCER) in Geneva, Switzerland and was test deployed through the Google Play Store in Switzerland during last year's influenza season. It is currently being prepared for a broader European deployment.

Author Keywords
influenza monitoring; participatory data collection; crowdsourcing; mobile phone sensors

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous; J.3 [Computer Applications]: Life and Medical Sciences - Health

Introduction

Crowdsourcing information from user's smart phone sensors has recently become a popular approach for monitoring various real world phenomena. Examples range from monitoring crowd motion at public events [11], road conditions [1] to social interactions [10]. Smart phone sensor data has also been used for monitoring personal health and health related behaviours. Beyond a variety of commercial mostly physical activity oriented monitoring applications varous diagnostic systems have been demonstrated in research. Thus, for example, our group, has previously shown how a combination of motion sensor data averaged over a day with mobility pattern analysis and communication pattern analysis can be used to diagnose depressive and manic episodes [13]. In this paper we build on such work and the successful use of participatory monitoring of Flu through self-reported symptoms to explore the use of crowd sourcing for detection of epidemic Flu spreading augmenting self-reported symptoms using mobile phone sensors.

Epidemiological Models and Data

Models and data characterizing contagious disease and the mechanisms via which it spreads have greatly enhanced the ability to 1) further scientific understanding of how disease spreads, 2) monitor disease (surveillance) and predict disease. One approach is statistical, collecting data to estimate incidence, prevalence and enable forecast using time series models [24]. A complimentary approach is to model the spread of the disease through individual level models and approximations thereof, which capture transmission dynamics at different levels of granularity [9, 5, 22]. These models are initialized and parametrized using biological characteristics of the pathogen and aspects of human behavior that affect transmission of the pathogens. The coarsest models are population level compartmental models that differentiate between individuals according to their epidemiological state (infectious, susceptible, immune, symptomatic) but otherwise assume that individuals are homogeneous and mix at random. On the other end of the spectrum, the most fine-grained models are agent based and consider individual level and environmental details that may affect the spreading process, such as behavior, intrinsic characteristics that may affect the immune system and environmental factors.

The traditional source of data for monitoring and predicting disease incidence with statistical population level models, is the sentinel system. However, these reports have a delay and may underestimate prevalence. An alternative approach which is gaining momentum is participatory monitoring, whereby participants voluntarily report symptoms associated with Influenza-Like-Illness (ILI), and other epidemiologically relevant participant information. Combining this data with that obtained using traditional systems has yielded improved predictions of Flu, given that the data sources capture different information [24].

Transmission based models are parametrized with data that characterizes the multi-scale structure of human mobility, form long range travel to measurements of proximity between individuals to establish the possibility of direct transmission. Various different proxies for human mobility have been used to model epidemic spreading [28]. One of the earliest approaches was to couple population centers using human traffic flows between them, quantified e.g. by airline travel [17, 8, 31, 6] or other commuting data to calibrate Meta-population models [2, 4]. More fine grained mobility traces can be inferred from GPS mobile sensors [12].

Proximity and face-to-face contact between individuals is typically modeled through the abstraction of contact networks [21, 18, 9]. There are multiple different approaches for inferring and measuring these interactions. For example, Radio Frequency Identification Devices (RFID) yield precise measurements of proximity [7]. However, Bluetooth sensors are also a a useful proxy for proximity and and physical contact between individuals [27, 26]. Initially the focus was on characterizing the heterogeneity in the contact and activity patterns across different individuals [23, 20]. However, there is clear evidence that both the number and duration of contacts have an effect on the transmission of disease [25]. In fact, recent work highlights the importance of temporal structure of contacts in understanding spreading processes [16, 15, 30].

Next Generation Disease Monitoring and Forecasting
It is clear that mobility and contact data used for modeling disease spread is most useful when it captures multiple spacial scales, heterogeneity between individuals and has a fine temporal resolution. Here we report on the design and early results of the grippeNET App, which is the first mobile application for participatory Flu monitoring. Starting from InfluenzaNet [19], a European participatory network of web platforms for the monitoring of ILI, grippeNET has developed and implemented a new paradigm of data collection, augmenting the survey data that users contribute with data collected through mobile phone sensors. The real-time data collected has a high temporal and spatial granularity, is collected over a long period of time and over a large and widespread population. This provides data well suited to calibrate transmission models.

Beyond an extension and scaling up of data collection the innovation of our approach is two-fold. First, we collect sensor information on user mobility, contact and activity patterns in a privacy preserving fashion. We do this by computing epidemiologically relevant characteristics locally on the phone, transmitting only processed data that does not reveal coordinates or identities of individuals. This is of prime importance if the efforts to collect such data for the social good are to continue without violating individual's concerns for privacy. The second innovation of this system is that the combination of self-reported symptoms with the data collected form sensors enables the use of a new type of model. Namely, we can train machine learning models that use information both on an individual's characteristic behavior (and how it compares to the behavior of other individuals) and also information on how an individual deviates from their characteristic behavior. Prior investigations on a small cohort for limited time indicate that characteristics of individual mobility can be predictive of Flu incidence [3]. These models can improve monitoring and prediction of infection through estimation of flu transmission at an individual level. Individualized risk estimates open up the potential to better inform individuals so that they can voluntarily modify their behavior and minimize their risk of becoming ill or transmitting the disease [14].

The grippeNET Platform
The grippeNET App is currently publicly available in the Google Play Store for monitoring, analyzing and mapping the spread of influenza (currently in Switzerland, to be extended EU-wide in future). The App offers the same functionality as the original Swiss web based service but is more user-friendly, and interactive. Reminders to users are sent by notifications. The App allows participants to fill in weekly flu surveys from their smart devices and to add to a collection of additional data collected through the device sensors aimed at enhancing the epidemiological studies. In return, users gain access to timely and comprehensive insights into the data and the analyses. Thus, users are not just data providers but also users of the system. In future versions, we can provide a forecast for flu based on the data generated by the users, in combination with the data collected by the sentinel network, in a weather-forecast-like

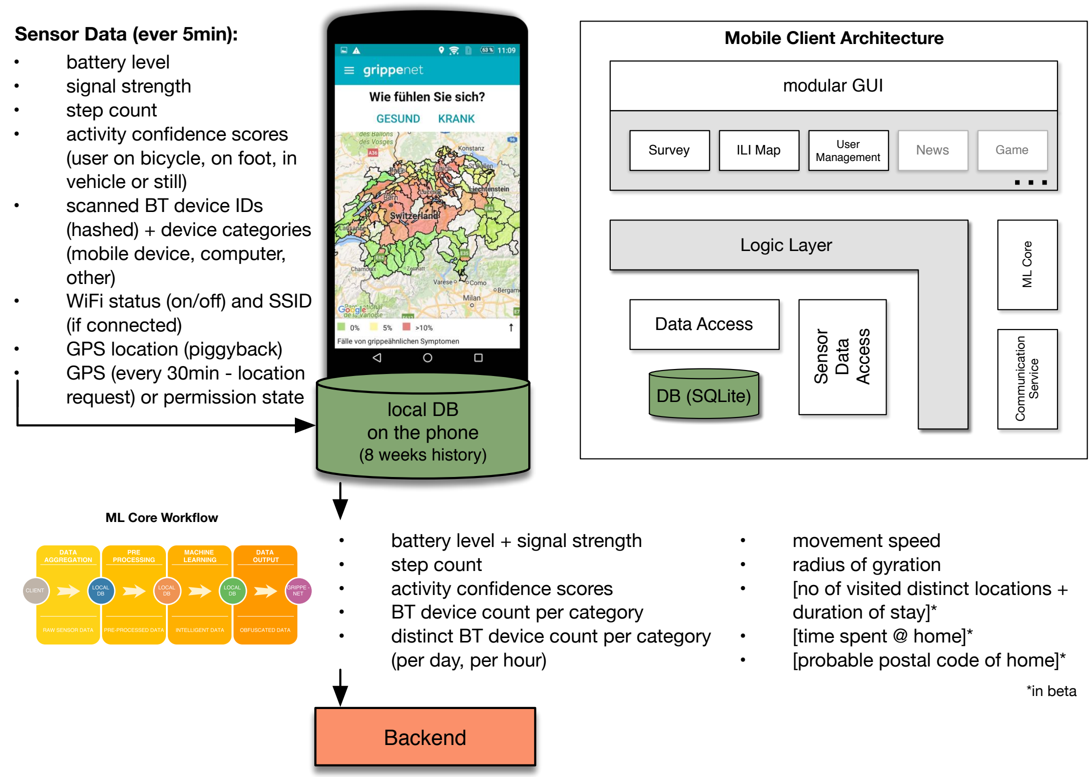

Figure 1: List of sensor data collected on the user's device, the App client's modular architecture, the machine learning core workflow and a list of aggregated data sent to the backend.

application.

Ethical and User Oriented Design
One of the key innovations of the grippeNET App is to build a system where users are motivated to participate in influenza monitoring, and continue to participate, in contributing their data for the social good. Our approach is built around engaging users, by giving them back aggregated information they will find interesting and useful while at the same time protecting their privacy. The grippeNET App has already served as a platform for studies in the field of human-computer interaction to examine what factors play a role in motivating a person to participate [29].

The modular nature of the App's code can be seen in Figure 1 and makes it possible to provide extension with minimal effort. As proof of concept, prototypes of a gamification and an extended statistics module connecting to the Fitbit API have been implemented to further provide value to the user and explore possibilities for an extended study. A first version of the gamification feature has left beta stadium and was released with the latest version of the App.

Moreover, we designed an extensive research study integrating data collected from the App and the web platform to continue the traditional monitoring analyses of participatory monitoring systems and also carry our novel predictive analytics using state of the art machine learning methods. We submitted the study for approval to the Regional Research Ethics Committee (CCER) in Geneva, Switzerland. Our approach was to write a privacy policy that follows the requirements and procedures of informed consent form for participation in a medical study. Our application was successful, and we were granted permission to operate "without reservation" from the committee, with a few optional recommendations currently being implemented.

Many agile design steps have been iterated, considering input from software engineers, epidemiologists, data scientists, human-computer-interaction and ethics experts ensuring sufficient data quality, stability, scalability and privacy protection.

The system uses state of the art encryption technologies, hashing, data reduction and aggregation algorithms on any bit of information before it leaves the user's device. Fulfilling all requirements for compatibility to the current InfluenzaNet system, the grippeNET backend consists of a special server separating smartphone-collected data from participant's personal data inside the InfluenzaNet user profile. The link is created via a unique ID to which researchers can't have access.

Data Collection
Sensor data is always recorded in the background using battery preserving methods and stored in a local database on the user's device. Special algorithms have been developed and are continuously improved for aggregating and obfuscating the raw data for privacy protection while still ensuring sufficient data quality. Figure 1 shows an overview over the collected data and calculated features by the grippeNET App.

To gain location-based features, we are taking usage of the GPS sensors integrated in most modern smart phones to sample a location trace which consists of the latitude, longitude and time stamp. The locations are sampled irregularly due to battery saving and fluctuation in user's permission or phone usage and stored in the device's local database. The database is not exposed to any 3rd party library or server, in the interest of privacy protection. We calculate simple location based features as well as more complex ones, utilising the ML core.

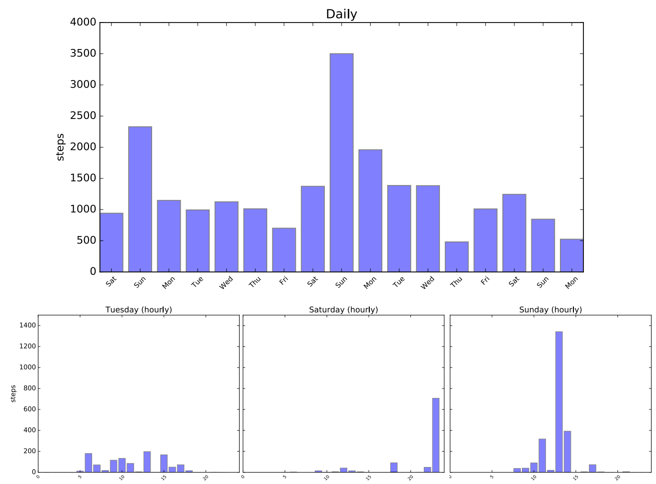

Figure 2: Daily step count over two weeks and hourly step count from a Tuesday, Saturday and a Sunday of a full time employed male person, 60 year old.

Data Examples

Figure 2 shows the number of steps calculated by the Android OS based on a device internal IMU (inertial measurement unit) from one person's phone over the course of two weeks. The daily step count during the first two weekends hint at this person seems to be more active on Sundays than during the rest of the week. Three charts compare a Tuesday, a Saturday and a Sunday of the same person's step count on an hourly interval. It can be seen that the distribution of step count over one day can vary and might be used to retrieve not only a proxy for physical activity but also context information, especially fused with other sensor data collected by the App.

Battery power level of the mobile device, we think looks like

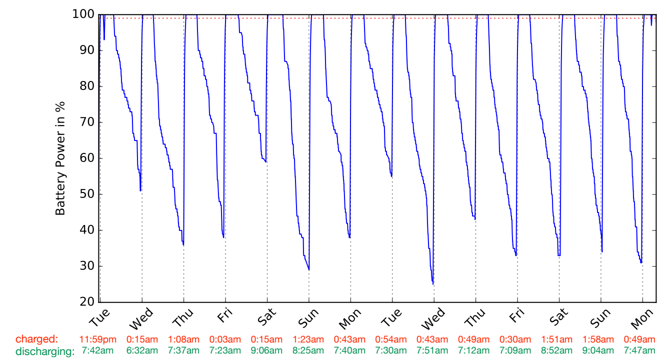

Figure 3: Battery power level over two weeks from a device of a full time employed male individual, 35 years. The grid lines mark each days 12am time. One charge and discharge time based on a 99% threshold (red, dotted line) can be found below the time axis.

a promising feature containing information about a user's behaviour. In Figure 3 it can be seen that using a basic one-threshold mechanic we might have an idea about the user's sleep cycle and daily routine. During the plotted two weeks, this user did never charge his phone during the day and his battery has a rest charge between 30% and 60% in the evening. The discharge times show when the user starts using his phone away from the charger, eg. when starting the day. Charge times should everyday be approximately equidistant to the times when the user plugged the charger into the phone, eg. before going to bed. If we would interpret the time between charge and discharge times as sleep time, it would mean that this user sleeps a longer on weekends than during the week.

In Figure 4 the number of different Bluetooth IDs scanned every 5min is plotted over one week for two random users. It is interesting to see that the profile for both looks very different. User A had the Bluetooth transceiver switched off

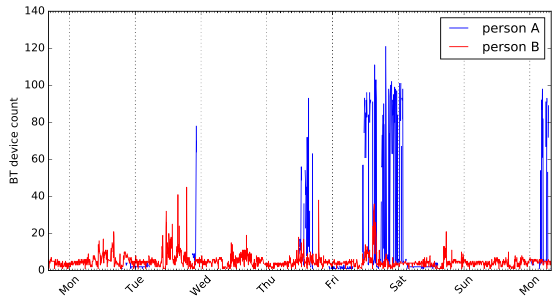

Figure 4: Total bluetooth scan device count in 5min intervals over one week comparing two random people. The grid lines mark each days 12am time.

most of the time. When activated either very few devices were seen or a very high amount of devices. Crowds of people or buildings with a lot of technology or Bluetooth beacons for localization could lead to such a high number of devices seen. For user B Bluetooth scans were possible continuously. It can be seen that the maximum of seen devices is a lot lower than most of the scans from User A. There are multiple spikes in the middle of the day during working days and a relatively flat curve on Sunday. Friday and Saturday evening more devices have been scanned then during other evenings. We think that with more sophisticated algorithms on the relatively high time resolution combined with the category of seen devices will lead to useful features for enriching disease spreading models.

Conclusion

The key lesson learned from the work so far has been the understanding of the sensing and analysis requirements and constraints in the triangle between (1) the information requirements of relevant epidemiological models (2) ethical, privacy and user acceptance considerations and (3) the technical feasibility of extracting (1) under constant challenge of (3) while observing (2). In the current App version we have implemented those collection and analysis strategies in a modular, easily extendable way. We have also obtained and examined initial data from a first real life deployment. The integration of the App with the existing InfluenzaNet platform has been crucial to getting practitioners from the field of epidemiology on board. At the same time it has been the source of a whole range of practical implementation issues.

Acknowledgements

This work is partially funded by the European Community's H2020 Program under the funding scheme "FETPROACT-1-2014: Global Systems Science (GSS)", grant agreement #641191 CIMPLEX.

REFERENCES

1. G Alessandroni, LC Klopfenstein, S Delpriori, M Dromedari, G Luchetti, B Paolini, A Seraghiti, E Lattanzi, V Freschi, A Carini, and others. 2014. Smartroadsense: Collaborative road surface condition monitoring. *Proc. of UBICOMM-2014. IARIA* (2014).

2. Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. 2009. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences* 106, 51 (2009).

3. Gianni Barlacchi, Christos Perentis, Abhinav Mehrotra, Mirco Musolesi, and Bruno Lepri. 2017. Are you getting sick? Predicting influenza-like symptoms using human mobility behaviors. *EPJ Data Science* 6, 1 (2017).

4. Vitaly Belik, Theo Geisel, and Dirk Brockmann. 2011. Natural Human Mobility Patterns and Spatial Spread of

Infectious Diseases. *Phys. Rev. X* 1 (2011). Issue 1.

5. Fred Brauer, Carlos Castillo-Chavez, and Carlos Castillo-Chavez. 2012. *Mathematical models in population biology and epidemiology*. Vol. 1. Springer.
6. Dirk Brockmann and Dirk Helbing. 2013. The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. *Science* (2013).
7. Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. 2010. Dynamics of person-to-person interactions from distributed RFID sensor networks. *PloS one* 5, 7 (2010).
8. Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. 2006. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the USA* 103, 7 (2006).
9. Odo Diekmann and Johan Andre Peter Heesterbeek. 2000. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Vol. 5. John Wiley & Sons.
10. Nathan Eagle and Alex Pentland. 2006. Reality mining: sensing complex social systems. *Personal and ubiquitous computing* 10, 4 (2006).
11. Tobias Franke, Paul Lukowicz, and Ulf Blanke. 2015. Smart crowds in smart cities: real life, city scale deployments of a smartphone based participatory crowd management platform. *Journal of Internet Services and Applications* 6, 1 (2015).
12. Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. 2008. Understanding individual human mobility patterns. *nature* 453, 7196 (2008).
13. Agnes Gruenerbl, Amir Muaremi, Venet Osmani, Gernot Bahle, Stefan Oehler, Gerhard Troester, Oscar Mayora, Christian Haring, and Paul Lukowicz. 2015. Smartphone-Based Recognition of States and State Changes in Bipolar Disorder Patients. *IEEE Journal of Biomedical and Health Informatics (JBHI)* 19 (2015).
14. Dirk Helbing, Dirk Brockmann, Thomas Chadefaux, Karsten Donnay, Ulf Blanke, Olivia Woolley-Meza, Mehdi Moussaid, Anders Johansson, Jens Krause, Sebastian Schutte, and others. 2015. Saving human lives: What complexity science and information systems can contribute. *Journal of statistical physics* 158, 3 (2015).
15. Petter Holme and Naoki Masuda. 2015. The basic reproduction number as a predictor for epidemic outbreaks in temporal networks. *PloS one* 10, 3 (2015).
16. Petter Holme and Jari Saramaeki. 2012. Temporal networks. *Physics Reports* 519, 3 (2012).
17. L. Hufnagel, D. Brockmann, and T. Geisel. 2004. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America* 101, 42 (2004).
18. Matt J Keeling and Ken TD Eames. 2005. Networks and epidemic models. *Journal of the Royal Society Interface* 2, 4 (2005).

19. E. Carl Koppeschaar, Vittoria Colizza, Caroline Guerrisi, Clément Turbelin, Jim Duggan, John W. Edmunds, Charlotte Kjelsø, Ricardo Mexia, Yamir Moreno, Sandro Meloni, Daniela Paolotti, Daniela Perrotta, Edward van Straten, and O. Ana Franco. 2017. Influenzanet: Citizens Among 10 Countries Collaborating to Monitor Influenza in Europe. *JMIR Public Health Surveill* 3, 3 (2017).

20. Yamir Moreno, Romualdo Pastor-Satorras, and Alessandro Vespignani. 2002. Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B-Condensed Matter and Complex Systems* 26, 4 (2002).

21. Mark EJ Newman. 2002. Spread of epidemic disease on networks. *Physical review E* 66, 1 (2002).

22. Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. 2015. Epidemic processes in complex networks. *Reviews of modern physics* 87, 3 (2015).

23. Romualdo Pastor-Satorras and Alessandro Vespignani. 2001. Epidemic spreading in scale-free networks. *Physical review letters* 86, 14 (2001).

24. Daniela Perrotta, Michele Tizzoni, and Daniela Paolotti. 2017. Using Participatory Web-based Surveillance Data to Improve Seasonal Influenza Forecasting in Italy. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland.

25. Piero Poletti, Roberto Visintainer, Bruno Lepri, and Stefano Merler. 2017. The interplay between individual social behavior and clinical symptoms in small clustered groups. *BMC Infectious Diseases* 17, 1 (2017).

26. Vedran Sekara and Sune Lehmann. 2014. The strength of friendship ties in proximity sensor data. *PloS one* 9, 7 (2014).

27. Arkadiusz Stopczynski, Vedran Sekara, Piotr Sapiezynski, Andrea Cuttone, Mette My Madsen, Jakob Eg Larsen, and Sune Lehmann. 2014. Measuring large-scale social networks with high resolution. *PloS one* 9, 4 (2014).

28. Michele Tizzoni, Paolo Bajardi, Adeline Decuyper, Guillaume Kon Kam King, Christian M Schneider, Vincent Blondel, Zbigniew Smoreda, Marta C González, and Vittoria Colizza. 2014. On the use of human mobility proxies for modeling epidemics. *PLoS computational biology* 10, 7 (2014).

29. Patrick Tobien, Lars Lischke, Marco Hirsch, Robert Krüger, Paul Lukowicz, and Albrecht Schmidt. 2016. Engaging People to Participate in Data Collection. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. ACM, New York, NY, USA.

30. Eugenio Valdano, Michele Re Fiorentin, Chiara Poletto, and Vittoria Colizza. 2018. Epidemic Threshold in Continuous-Time Evolving Networks. *Physical Review Letters* 120, 6 (2018).

31. Olivia Woolley-Meza, Christian Thiemann, Daniel Grady, Jake Jungbin Lee, Hanno Seebens, Bernd Blasius, and Dirk Brockmann. 2011. Complexity in human transportation networks: A comparative analysis of worldwide air transportation and global cargo ship movements. *European Physical Journal B* 84, 4 (2011).