UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Quantifying agent impacts on contact sequences in social interactions

Dekker, M.M.; Blanken, T.F.; Dablander, F.; Ou, J.; Borsboom, D.; Panja, D.

DOI
10.1038/s41598-022-07384-0

Publication date
2022

Document Version
Final published version

Published in
Scientific Reports

License
CC BY

Link to publication

Citation for published version (APA):
Dekker, M. M., Blanken, T. F., Dablander, F., Ou, J., Borsboom, D., & Panja, D. (2022). Quantifying agent impacts on contact sequences in social interactions. *Scientific Reports*, *12*, Article 3483. https://doi.org/10.1038/s41598-022-07384-0

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:15 Oct 2026

scientific reports

Check for updates

OPEN

Quantifying agent impacts on contact sequences in social interactions

Mark M. Dekker[1,2]✉, Tessa F. Blanken[3], Fabian Dablander[3], Jiamin Ou[1,4], Denny Borsboom[3] & Debabrata Panja[1,2]

Human social behavior plays a crucial role in how pathogens like SARS-CoV-2 or fake news spread in a population. Social interactions determine the contact network among individuals, while spreading, requiring individual-to-individual transmission, takes place on top of the network. Studying the topological aspects of a contact network, therefore, not only has the potential of leading to valuable insights into how the behavior of individuals impacts spreading phenomena, but it may also open up possibilities for devising effective behavioral interventions. Because of the *temporal* nature of interactions—since the topology of the network, containing who is in contact with whom, when, for how long, and in which precise sequence, varies (rapidly) in time—analyzing them requires developing network methods and metrics that respect temporal variability, in contrast to those developed for static (i.e., time-invariant) networks. Here, by means of event mapping, we propose a method to quantify how quickly agents mingle by transforming temporal network data of agent contacts. We define a novel measure called *contact sequence centrality*, which quantifies the impact of an individual on the contact sequences, reflecting the individual's behavioral potential for spreading. Comparing *contact sequence centrality* across agents allows for ranking the impact of agents and identifying potential 'behavioral super-spreaders'. The method is applied to social interaction data collected at an art fair in Amsterdam. We relate the measure to the existing network metrics, both temporal and static, and find that (mostly at longer time scales) traditional metrics lose their resemblance to *contact sequence centrality*. Our work highlights the importance of accounting for the sequential nature of contacts when analyzing social interactions.

Human behavior plays a central role in generating patterns of interaction that allow for the spreading of a great variety of entities—from rumors[1] to viruses[2] to memes[3]. Most recently, the importance of such patterns of interaction has been borne out by the COVID-19 crisis. A virus like SARS-CoV-2, which caused the COVID-19 pandemic, spreads through physical contact, or through aerosols that have a finite range; for the virus to be transmitted, people will have to have make close contacts with each other. It is for this reason that behavioral interventions, aiming to break the chains of close human contacts, were our only weapon against COVID-19 for the better part of 2020 and remain so until vaccine coverage is sufficient.

As such, the pandemic underscored the need for better models and measures of human behavior and the contact patterns it generates. On the technological side, recording individual-level contacts through wearable sensor technology has recently seen considerable development that render the study of contact patterns scientifically feasible[4–6]. For example, albeit still imperfect, sensor accuracy is beginning to approach the level and scale needed to inform epidemiology and to model interventions[7–9]. In contrast, on the methodological side, we still lack adequate tools to assess, represent, and model the myriad ways in which human behavior generates patterns of individual-level contacts. This limits progress in many disciplines that require such information, but especially hampers progress along the behavior-epidemiology interface that is so important in improving our preparedness to deal with the current and future pandemics. This paper develops methodologies that can begin to fill this gap, by exploiting connections among improved empirical assessments of contact patterns and novel quantitative metrics that are sufficiently advanced to analyze behavioral contact patterns. In particular, it focuses

[1]Department of Information and Computing Sciences, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands. [2]Centre for Complex Systems Studies, Utrecht University, Minnaertgebouw, Leuvenlaan 4, 3584 CE Utrecht, The Netherlands. [3]Department of Psychological Methods, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 VZ Amsterdam, The Netherlands. [4]Department of Sociology, Utrecht University, Padualaan 14, 3584 CH Utrecht, The Netherlands. ✉email: m.m.dekker@uu.nl

on quantifying the role of each individual in the overall contact patterns, reflecting the agent's impact on the potential of (e.g., epidemiological) spreading in a system.

To achieve these goals, we utilize a network approach in which patterns of contacts which are generated by individual behavior are mapped on to a contact network. In such a network, each individual corresponds to a node and a link between two nodes represents a contact between two individuals. The advantage of using a network approach is that a general spreading process of some entity (of which pathogen spreading in epidemiology is one, where the concerned pathogen is the entity) can be analyzed by means of a two-level description—the "infrastructure", and the process on top of it: (a) a contact network, borne out of people's behavior in social interactions that constitutes the infrastructure, and (b) a pathogen spreading process occurring within a population, which then becomes a (dynamical) process that uses this infrastructure as pathways to transmit from an infected to a susceptible agent. From a network science perspective, it then becomes natural to expect that the topological properties of the interaction networks—the infrastructure—can be used to analyze the *potential* of spreading dynamics. In addition, this approach allows us to analyze the relative importance of different individuals' potential of spreading by analyzing the effect that removing the individual would have on the network structure. Individuals who contribute greatly to the topological aspects of the network may then represent potential targets of intervention. (In contrast, social distancing is a non-pharmaceutical intervention measure against SARS-CoV-2 transmission by means of globally disrupting the infrastructure).

It is important to note, however, that spreading dynamics can take on many forms: infectious diseases will spread differently from rumors or political opinions, which are yet different from spreading of delays in network transport[10]. Nevertheless, the common element they all share is (deterministic or stochastic) transmission upon contact, and spreading is therefore dependent on the topology of the contact network: who interacts with whom, for how long and in which precise sequence. For example, if a person *A* spreads the rumor, a person *B* can only spread the rumor to person *C* if (s)he has been in contact with *A* before the contact with *C*. Tracking these sequences of interactions, therefore, provides the basis of any spreading phenomenon. While the work presented here was developed in the context of modeling behavior during the COVID-19 pandemic, the importance of better models for human behavior in generating contact networks should therefore be seen as generic.

Before a satisfactory analysis of this process can be found, however, several problems that arise in the currently used network analytic techniques (for evaluating the structure of contact networks and the role of different individuals in generating them) must be addressed. First and foremost, in reality, a contact network is dynamic, i.e., the topology of a contact network is changes over time. In Network Science such networks are called *temporal networks*. Many existing methods that analyze contact networks, empirically obtained through wearable sensors[11], treat contact networks as static, since, for instance, recorded contacts may not be adequately logged or the analysis needs to be simplified by collapsing networks over the time domain. From a Network Science perspective, this is in fact reminiscent of one way to reduce the complexity of a temporal network by downscaling it to a static one: a network with a time-invariant topology, obtained through aggregation of all edges over a certain time interval[6,7,12–14]. However, information on who interacts with whom, for how long and in which precise sequence gets lost in such aggregations. Static representations therefore do not reflect the dynamic aspects of human behavior: changing interactions, structures and sequences[4,5,12,15–19] have been shown to have large impacts on various dynamical processes[16,20–22]. Static representations miss a crucial opportunity to intervene at the contact network level, e.g., by targeting interventions at specific points in the time domain. Secondly, while there are *static* topological metrics available to quantify the impact of agents on spreading mechanisms, static network techniques cannot differentiate between agents whose contacts are structured in the same way but located at different time points. Examples of such topological metrics are degree, closeness or betweenness centrality[23,24], *k*-shell decomposition algorithm[25], methods based on clustering and the topology potentials[26], improved coreness centrality and eigenvector centrality[27], influential individuals[28].

More prudent (existing) approaches to deal with the complexity of temporal contact networks entail setting a threshold to filter out the non-essential edges or edges that exist only by chance: for example, Grabowicz *et al.* used a simple threshold based on the number of events (i.e., count of edges across time) between two nodes[29]; Kobayashi defined a temporal null model to identify pairs of nodes having more interactions than expected given their activities[30]; and in yet another thread of study, Mellor *et al* introduced the temporal event graph (TEG), which uses events (interactions between two individuals) as nodes and shared event attendees as edges in a directed graph[31,32]. Some of these efforts focus on only part of the complexity of temporal networks and provide possible ways to identify structures[29,30], communities[4,22,33] and quantify connectivity[32,34,35]. Other related papers on the topic discuss temporal motifs[36] and percolation in light of weighted temporal event graphs[37], or propose an event embedding technique to obtain a low-dimensional representation of temporal networks[35]. However, in the context of spreading phenomena, these metrics are not meant for quantifying agent impacts on contact chains in human interactions (and thereby the spread of a concerned entity). We also use event mapping, which we explain in "Methods" section.

A more principled approach to unraveling behavioral heterogeneity and the impact of individual actors, which we develop in this paper, is to consider the full dynamics of people's contacts in the form of a temporal network. Specifically, we provide a way forward by developing (a) a systematic methodology for representing people's (dynamical) contact networks as event graphs, and (b) techniques to analyze the role that individuals play in generating these network structures. The latter information can be used to assess individuals' impacts on the contact sequence *while respecting (temporal) sequence of contacts*.

This paper is structured as follows. We start by introducing the methodology to extract an individual's impact on contact sequences in a system in "Methods" section, for which we need to introduce a reconstruction of the temporal graph using 'event mapping'. In "Application to human interaction data" section, we apply the methodology to data collected during an art fair in Amsterdam in August 2020, where we also provide experimental details and discuss the evolution of the most important metrics over time, as well as relate them to the existing network

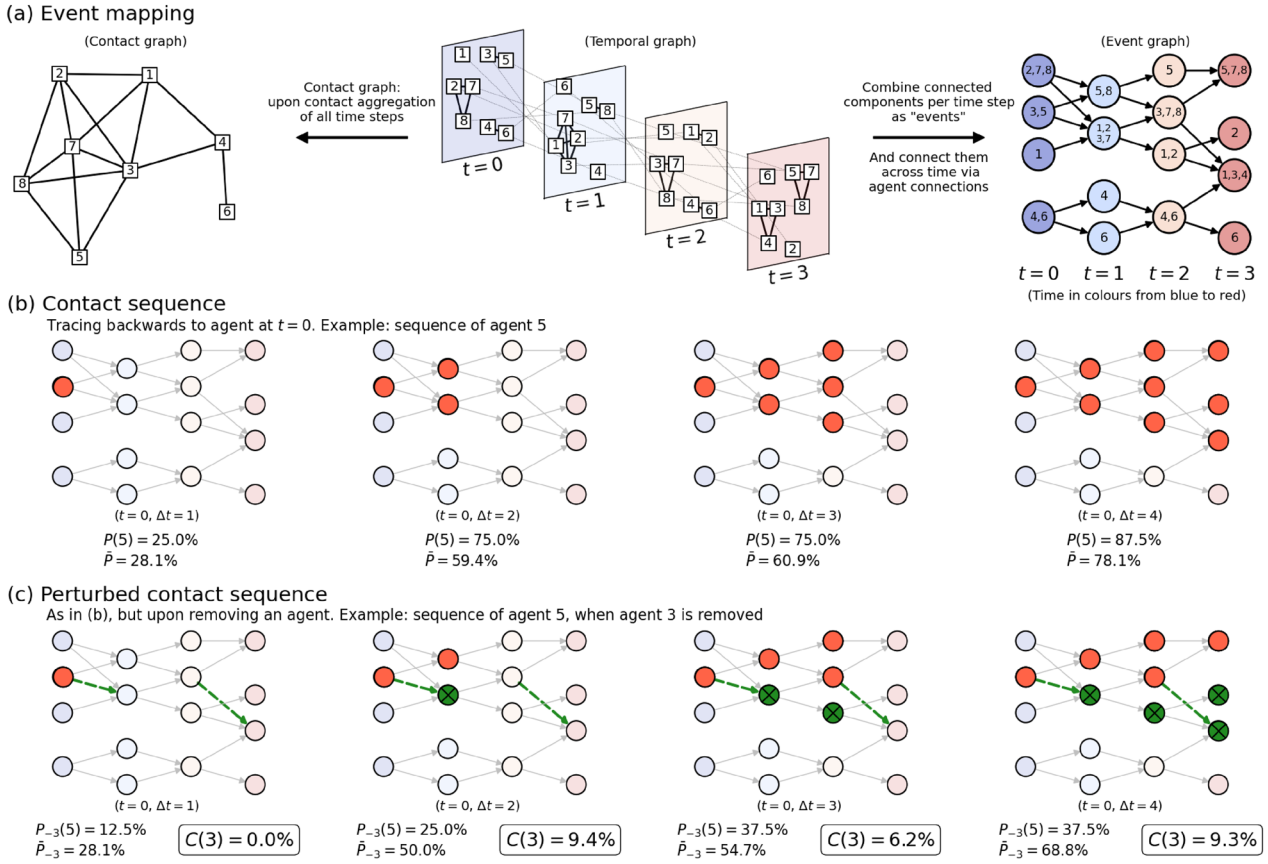

Figure 1. Illustration of the event mapping and an individual's contact sequence centrality using a toy example. All throughout this figure, individual agents are in squares (and numbers), events are in circles. **Panel (a)**: In a temporal network, the time evolution of the contact graph (middle) captured in a sequence of snapshots. Contact data over the four snapshots can be aggregated to obtain a static (and aggregated) contact network (left), but in doing so, the sequential information is lost. We term the connected network components within each snapshot 'events', they are seen as nodes in an event graph (right). In an event graph the event-to-event trajectories of agents are the links. Colors of circles and snapshots indicate time from blue to red. **Panel (b)**: Example of a contact sequence when tracing all events and subsequent agents back to agent 5 at $t = 0$. Starting time t, time elapsed Δt, the percentage unique agents $[P(5)]$, traceable to 5 and the average percentage of agents in the contact sequence $\bar{P}$ are indicated. **Panel (c)**: Same as (**b**), but after removal of agent 3, resulting in a 'perturbed' contact sequence of agent 5. The (now absent) links that were associated with agent 3 are in dashed green arrows, and the events that are consequentially not traceable to 5 anymore are colored green. The associated (lowered value of) $P_{-3}(5)$ is shown. Averaging over all agents provides $\bar{P}_{-3}$, and allows us to calculate the *Contact Sequence Centrality* of agent 3: $C(3) = \bar{P} - \bar{P}_{-3}$, which is clearly marked in the bottom panel. For abbreviation, we excluded the t and Δt between brackets when showing values of P, $\bar{P}$ and C.

metrics like contact degree. In "Discussion and conclusion" section, we discuss these findings and relate them to the broader discussion on spreading phenomena in networked systems and end with a few concluding remarks.

Methods

In this section, we introduce the mapping of temporal network data into an event graph in order to facilitate the identification of agent impacts on a sequence of contacts.

Event mapping of contact sequences. We track the contact sequences of the agents in the following manner. In Fig. 1a we show an example of a (randomly generated) temporal system with eight agents and four time steps $t \in \{0, 1, 2, 3\}$ in three types of graphs. In the center, the temporal graph is shown where agents are denoted in squares. The contacts among the agents at every snapshot are denoted by black links, while the agents' movements across snapshots are shown by gray links. To the left of the panel, we show the corresponding aggregated *contact network* over the entire time range, wherein two agents are connected by an edge if they have been in contact at any of the four time steps: the aggregated contact network therefore does not distinguish any sequence between the interactions.

To the right of Fig. 1a, we combine connected components within time snapshots into single circular nodes, which we refer to as *events*[38], resulting in yet another graph: the *event graph*. We use a *non-weighted* version of event mapping, where links between events are not weighted by the number of crossing participants. Events

may consist of either single or more agents, as marked by the numbers indicated inside each node. The events are linked by the 're-usage' of agents; this is exactly how every agent can act like a spreader between consecutive events, which makes event mapping a natural choice for describing the dynamical evolution of a spreading phenomenon. For example, at $t = 0$, agent 5 is interacting with agent 3. From there, they go their separate ways: at $t = 1$ agent 5 interacts with agent 8, and agent 3 interacts (in a larger group) with agents 1, 2 and 7. Indeed, the event (3, 5) at $t = 0$ links to both (5, 8) and (1, 2, 3, 7) at $t = 1$. Using these events, one can convert the (temporal) network snapshots to an event graph (directed in time), shown at the right of Fig. 1a. Not only does this *event mapping* collapse complex temporal network data into a single directed graph, but most importantly, it also preserves the sequences of contacts. [An important question is to what extent the information for spreading phenomena is already embedded in the static contact graph, and when it is important to do event mapping. This is a question we address multiple times in this paper, by comparing the results of our metrics (defined below) to properties of the static contact network.].

The event graph allows for an intuitive analysis of which (agent contacts in) events in later time steps are eventually traceable back to the an event in earlier time steps via both direct and indirect connections. An example of such a contact sequence is given in Fig. 1b. There, we track the events that are linked to agent 5 at $t = 0$ (colored red), over four sequential time snapshots. At $t = 0$, only the event where agent 5 itself is involved is colored red [i.e., agents (3,5)]. Over time, 'secondary' events become colored red in the event graph, meaning that in due course of time, progressively more events, and the agents contained therein, can be traced backwards to agent 5 at $t = 0$. In the end, for most applications, how many agents can be traced backwards to agent 5 at $t = 0$ is the most relevant quantity (rather than the number of events, i.e., the number of red dots). This prompts us to define $P(j, t, \Delta t)$:

$$P(j,t,\Delta t) = \frac{\text{Number of agents traceable to agent } j \text{ at time } t \text{ within time interval } [t, t+\Delta t]}{N}, \tag{1}$$

where N is the total amount of agents. In other words, $P(j, t, \Delta t)$ is the fraction of *unique* agents at time $t + \Delta t$ that can ultimately be linked backwards to agent j at time t—clearly, $P(j, t, \Delta t)$ is a function of the starting time t and elapsed time Δt. In Fig. 1b, we show $P(5, 0, \Delta t)$ for different values of Δt, using agent 5 as 'agent zero': at $\Delta t = 0$, $P(5, 0, 1) = \frac{2}{8} = 25\%$ since at $\Delta t = 0$ two (3 and 5 itself) out of eight agents are linked to 5 at $t = 0$. Similarly, $P(5, 0, 2) = \frac{6}{8} = 75\%$ since there are six agents (1, 2, 3, 5, 7 and 8) that can be traced backwards to agent 5 at $t = 0$, and so on. Clearly, a high value of $P(j, t, \Delta t)$ for low Δt would indicate fast mingling of the agents in the time interval $(t, t + \Delta t)$, making the system highly prone to spreading phenomena from the perspective of starting with agent j at $t = 0$. Conversely, a low value of $P(j, t, \Delta t)$ for high Δt would imply the opposite. It is obvious that for fixed t, $P(j, t, \Delta t)$ is a monotonically increasing function of Δt, while for fixed Δt and a different initialization time t, $P(j, t, \Delta t)$ may vary and therefore possibly result in lower values with increasing t.

In panel (b), we use agent 5 as a mere example to illustrate the calculation of $P(j = 5, t, \Delta t)$. A more representative quantity is obtained by taking the average over all agents j:

$$\bar{P}(t,\Delta t) = \frac{1}{N}\sum_{j}^{N} P(j,t,\Delta t) \tag{2}$$

the value of which is also indicated in Fig. 1b. The value of $\bar{P}$ says something about the (unperturbed) contact sequences: high values indicate that spreading to many agents within $(t, t + \Delta t)$ is likely, and vice versa. We aim to identify the impact of each agent on the overall contact sequences. We measure this by looking at the $\bar{P}$ value when we remove an agent—the intuition being that the resulting change in $\bar{P}$ reflects the impact of the removed agent. Again, we start with the example of the contact sequence of agent 5 (i.e, $P(5)$ instead of $\bar{P}$). Now, to assess the impact of *another* agent—agent 3—on agent 5's contact sequence, we *remove* this agent from the system and show what it does to agent 5's contact sequence in Fig. 1c. The removal results in the absence of important links (denoted in green dashed lines) and, consequently, in limiting the contact sequence's reach (i.e., excluding the green nodes). The proportion of agents in agent 5's contact sequence is now written as $P_{-3}(5, t, \Delta t)$, where the subscript $-i$ refers to the removal of agent i. In particular, we have $P_{-3}(5, 0, 2) = \frac{2}{8} = 25\%$, which is lower than $P(5, 0, 2)$, since the large event comprised of agents 1, 2, 7 and the removed 3 cannot be traced back to agent 5 at $t = 0$ anymore. More generally, for any agents i, j the relation $P_{-i}(j, t, \Delta t) \leq P(j, t, \Delta t)$ holds, because the contact sequence is either equal or shrunk by the removal of agent i.

The impact of an agent on the contact sequence is naturally quantified by the difference in $\bar{P}$ between the unperturbed [panel (b)] and the perturbed event graph. For the impact of agent 3, the latter is shown in panel (c). Averaging over the contact sequences of each agent j, we obtain an expression for the impact of an individual agent on the contact sequence which we refer to as *contact sequence centrality*:

$$\begin{aligned} C(i,t,\Delta t) &= \frac{1}{N}\left[\sum_{j} P(j,t,\Delta t) - \sum_{j} P_{-i}(j,t,\Delta t)\right] \\ &= \bar{P}(t,\Delta t) - \bar{P}_{-i}(t,\Delta t). \end{aligned} \tag{3}$$

A high $C(i, t, \Delta t)$ indicates that the removal of agent i decreases the fraction of agents in an average contact sequence sharply, which reflects an important role of i in connecting contact sequences. This way, $C(i, t, \Delta t)$ becomes an attribute of agent i. It might just reflect that agent i itself has had a lot of contacts in the interval $(t, t + \Delta t)$ (i.e., the degree of agent i in the aggregated contact graph is high), but that is not necessarily the case. It may also be that agent i serves as the intermediary between two larger 'bubbles' or temporal communities,

which will become disjointed if agent i is removed from the system. Example values of $C(3, t, \Delta t)$ are shown in Fig. 1c for the case of agent 3.

Other agent-based topological network metrics. It is important to assess whether the contact sequence centrality C merely reflects static properties of the network topology that could have been found more easily using traditional methods, or whether it provides us with new information regarding an agent's role in the system. This prompts us to compare C values to various other agent-based metrics that have been around in the literature. Some of them are derived from the contact network like the one shown Fig. 1a on the left, obtained by aggregating all the temporal network snapshots into one single, unweighted, undirected contact network within a certain (or the whole) time interval. The *degree* of an agent i equals the amount of links this agents has, i.e. the amount of unique agents i had been interacting with. Similarly, the *contact betweenness* is the betweenness centrality in the aggregated contact network; betweenness centrality is defined as the fraction of all shortest paths (i.e., between all unique pairs of agents in the aggregated network) that passes through an agent. As for event maps of temporal networks, the *number of events* at any time snapshot is the amount of events of sizes larger than 1 (i.e., 'non-individual events') the agent attends. The *average event size* determines the average amount of attendants of events in general.

Finally, for temporal networks, many non-static quantities have been developed, as variants of their static counterparts, in order to study the temporal structural properties of the system[39,40]. While they have been developed for degree and closeness as well[41], we choose to compare $\Delta\bar{P}$ to two types of betweenness other than the aforementioned contact betweenness. To explain the difference, we refer to the graphs shown in Fig. 1a. On the left, we see the contact graph, in which the contact betweenness (already mentioned above) is defined. On the right, we see an event graph, where the nodes are events. The betweenness—i.e. fraction of shortest paths going through them—of nodes in this graph is computed, and when taking the average of all events where any agent i is participating, we obtain the value of what we refer to as the *event betweenness* of this agent. Likewise, we can focus on the network in the middle: the temporal graph. Nodes here are agents at a specific time snapshot. The betweenness in this graph is computed, and then we again average over all nodes in this graph belonging to each agent individually, which results in what we refer to as the *temporal betweenness* of an agent. The event betweenness and temporal betweenness are related, but not necessarily the same, as finer network structures within events are not incorporated in the event betweenness (while they are in temporal betweenness), and the event betweenness calculates shortest paths between pairs of events rather than pairs of agents.

Because of computational limitations, we choose to calculate temporal betweenness and event betweenness in time intervals smaller than the full data sets (i.e., we take ten subsequent 4-min time intervals). This results in multiple values per agent (each belonging to one of these intervals), of which the average is reported in later usage of these metrics.

Application to human interaction data

In this section, we perform event mapping and contact sequence calculations on human interaction data measured at an art fair in Amsterdam. After describing the experiment details in "Experiment and sensor data" section, we discuss the resulting $\bar{P}$ and C values in "Evolution of $\bar{P}$ and C" section, and compare the latter to other metrics in "Relating C to existing network metrics" section.

Experiment and sensor data. In August 2020, after the first peak of COVID-19 infections was decaying in the Netherlands, right before the start of the second wave of infections, an art fair was organized in Amsterdam with the goal of assessing the effectiveness of physical distancing interventions. For an illustration of the art fair, see Fig. 2. During this three-day art fair, different interventions to promote physical distancing were implemented: walking directions, face masks, and a buzzer-notification (a buzzing sound when coming within 1.5m of another visitor). The three day art fair was split into eleven time slots and the aforementioned interventions varied across these time slots. Each visitor was asked to wear a sensor that recorded the distance to the sensors of other visitors within line of sight (max 30m) using ultra wide band technology. Whenever two sensors were within 1.5m of one another the opposing tag ID was registered and the contact was logged locally (at a 1-s resolution), which was sent to a database via access points, placed near the entrance of the art fair. Researchers from the University of Amsterdam collected the questionnaire and sensor data and have published this in previous work[11,42]. The ethics review board of the University of Amsterdam (2020-CP-12488) approved data collection, and all participants provided informed consent before participating. All personally identifiable information used to link the questionnaire and sensor data has been destroyed. All methods were performed in accordance with the relevant guidelines and regulations.

From time to time, in most experiments, the interaction registration was hampered by not synchronizing to the access points well, altering some of the time stamps. Because the time stamps are crucial to the analysis in this paper, we chose to limit the analysis to parts of the experiments that did not contain such gaps (see Appendix A for a supplementary figure on choosing the correct parts of the data).

Table 1 shows an overview of the experiments, their (filtered) duration, conditions and the resulting amount of agents. More details can be found in Ref.[42].

Evolution of $\bar{P}$ and C. In Fig. 3, in order to illustrate differences among the experiments, we show the time evolution of the system property $\bar{P}$ and the contact sequence centrality C (for specified agents) in three of these experiments: 7, 9 and 11. The conditions behind these experiments are quite different as shown in Table 1: experiment 11 contained no interventions at all, while direction and buzzer feedback interventions are differently applied in experiments 7 and 9. Importantly, we use the experiments to highlight several aspects of $\bar{P}$ and C.

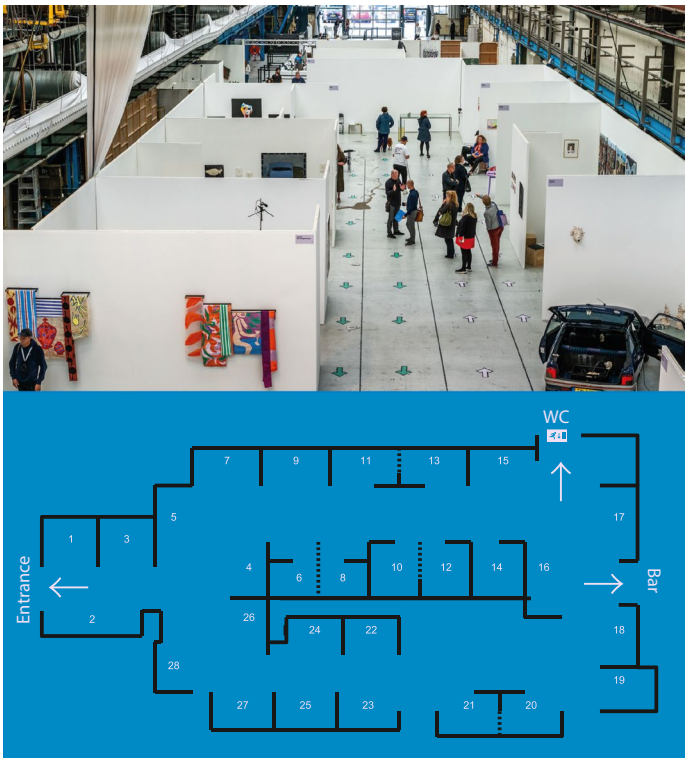

Figure 2. The art fair on Amsterdam on the first day with bidirectional walking directions (top) and schematic layout of the location (bottom). The art fair consisted of 28 stands, spanning 1080m^2. Visitors entered on the left of the plan in the bottom panel, where also the access point was placed. Figure adapted and reprinted with permission from Ref.[42].

Exp	Duration (min)	Walking direction	Supplementary intervention	SDS setting	# Agents	Avg. degree	Avg. degree per agent
Exp 2	19	Bidirectional	Facemask	No feedback	98	10.2	0.10
Exp 3	25	Bidirectional	None	No feedback	95	5.1	0.05
Exp 4	50	Bidirectional	Buzzer	Buzzer after 2 s	88	18.1	0.21
Exp 6	55	Unidirectional	None	No feedback	130	19.6	0.15
Exp 7	38	Unidirectional	Buzzer	Buzzer immediately stops after 2 s	88	7.9	0.09
Exp 9	85	No direction	Buzzer	Buzzer immediately stops after 2 s	109	17.3	0.16
Exp 10	42	No direction	Buzzer	Buzzer immediately persists after 2 s	120	11.7	0.10
Exp 11	44	No direction	None	No feedback	89	17.4	0.20

Table 1. Details of the conditions and attributes of the experiments. In the second column ('Duration') we note the filtered window durations, rather than the full experiment times. Event amounts and link frequencies are determined using only events with sizes larger than 1. Number of agents may differ from that in Ref.[42] because we focus on only a part of each experiment (see Fig. A.1). The average degree and the average degree per agent have been calculated from the aggregated contact network, aggregated over the full valid duration of the experiments.

First, in the top row, we show the evolution of $\bar{P}(t, \Delta t)$—see definition in Eq. (3)—as a function of (the starting time) t. A parameter that needs to be chosen is Δt, modulating the time interval across which the tracing between agents is done (see Fig. 1b for an example). This choice requires special attention, because while some agents may have most impact in the short run, others may have more impact on the long run. In other words: the individual differences in impact on contact sequences, the main topic of this paper, are dependent on the 'time scale' Δt we are interested in (we will come back to this later). In Fig. 3a–c, we choose $\Delta t = 300$ and 600 s. For example, the black curve represents the average fraction of unique agents in the time interval $(t, t + 300$

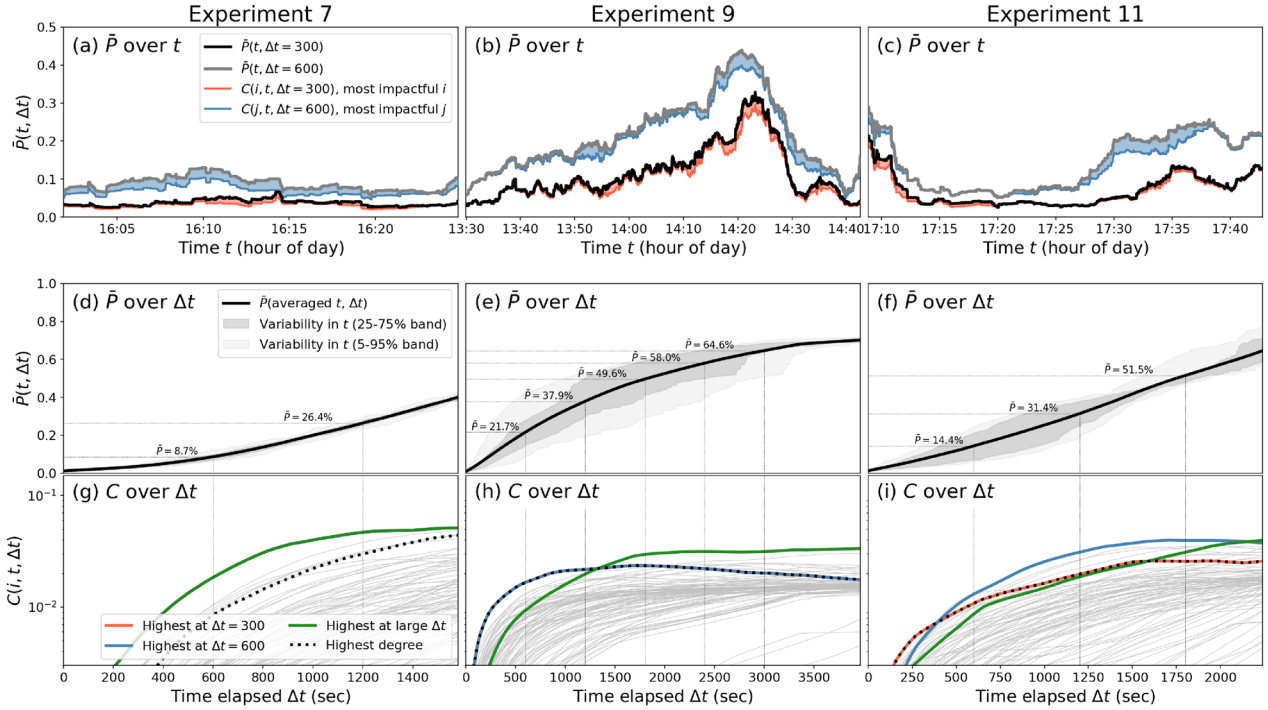

Figure 3. Evolutions of $\bar{P}$ and C of experiments 7 (left column), 9 (middle column) and 11 (right column). **Panels (a–c)**: Evolution of $\bar{P}(t, \Delta t)$ over time t at Δt =300 s (black/red) and Δt =600 s (gray/blue). We also show $\bar{P}_{-i}(t, \Delta t)$ (red and blue curve, respectively) and $C(i, t, \Delta t)$ (red and blue area) for the most impactful agents i at these levels of Δt. **Panels (d–f)**: Evolution of $\bar{P}(t, \Delta t)$ over elapsed time Δt, averaging over t, for the same experiments 7, 9 and 11. For a few values of Δt, we indicate the value of $\bar{P}(t, \Delta t)$. **Panels (g–i)**: Evolution of $C(i, t, \Delta t)$ for each agent i (in gray). Specific agents are denoted in colors: the agents with the highest C at Δt =300 (red) and 600 s (blue) as in panels (**a–c**), the agent with the highest C at a near-maximum Δt (i.e., 90% of the maximum time elapsed) in green, and the agent with the highest static contact degree as a black dashed line. Note that in panel (**g**), the red, blue and green lines overlap (only green shows), and in panel (**h**), the red and blue overlap (only blue shows).

s) in the contact sequence of any agent at time t. High values at time t (may) indicate that large and frequent events are taking place in the interval $(t, t + 300$ sec), involving many unique mingling people, as opposed to mingling only within specific bubbles or clusters. Clearly, the grey curve ($\Delta t = 600$) is always above the black curve ($\Delta t = 300$), which is expected, as over a longer time span $\bar{P}$ has more time to grow. Visual inspection of the panels (a)–(c) immediately reveals a number of differences in the evolution of $\bar{P}(t, \Delta t)$ among the different experimental settings: while experiment 7 remains rather flat at low values, experiment 9 contains a build-up towards much higher values (up to 0.30 for $\Delta t = 300$ sec), and experiment 11 lies in between the two and seems to be split in alternating phases of highs and lows. Without going into the specifics of these experiments, the shapes of the $\bar{P}$ curves can be interpreted as follows. A constant, flat curve (like in experiment 7) indicates that the interaction configuration is such that the potential of any spreading phenomenon does not vary much in time: if a disease or rumor would start at the start of the time interval, or in the middle, it would spread with roughly equal speed. However, a lot of variation in $\bar{P}$ (like in experiment 9) indicates that the structure of interactions and the temporal graph itself is different when starting at different time points. Sometimes, it promotes tracing many agents through contact sequences (e.g., at 14:20, we have that $\bar{P} = 40\%$ in 10 minutes) by means of very frequent events with randomized attendees, which may imply that the system is vulnerable to spreading dynamics like disease transmissions and rumor spreading. In other cases, the temporal network structure only contains narrow contact sequences, tracing only few agents to each other (e.g., at 13:30, we have that $\bar{P} = 5\%$ in 10 minutes), which is the case in a well-segregated agent population.

In Fig. 3a–c, to mark the effect of individual agents in the system, we also plot C after fictitiously removing the 'most impactful agent' at both levels of Δt plotted here: 300 and 600 s, in red and blue, respectively. Note that there are different ways to define the 'most impactful agent', and the correct choice is depending on the application. In particular, the ranking of agents in terms of their impact on the contact sequences is in reality dependent on both t and Δt, i.e., who is 'most impactful agent' varies over time. For illustration purposes we choose to average all C values over t and look at the agent with the highest value of C at fixed, chosen values of Δt (in other panels we use the same procedure). In other words, our definition of 'most impactful agent' is the agent that at a chosen value of Δt, has the highest value of C, averaged over the possible time starting points t. In the panels (a)–(c), the reduction in $\bar{P}$ (marked by the red and blue areas) if we would remove these high-impact agents remains relatively invariant over time t in experiment 7, while in experiments 9 and 11, there are clear moments in time where the most impactful agents seem to have their highest impact: for experiment 9, this is

near the peak, and for experiment 11, most impact is found either at the start, or around 17:27-17:37. The agents denoted in red and blue are not necessarily the same agent (see below).

While in the upper row, we see how $\bar{P}$ evolves over time, indicating where contact sequences promoted any spreading dynamics most, we now look at how $\bar{P}$ grows with Δt in panels (d)–(f), which reflects how many people you can trace via contact sequences within a given time interval size Δt. These are monotonically increasing curves: taking a larger time window allows the system reach equal or more agents, but never fewer. Cross-experimental differences can be observed in the slope at which $\bar{P}$ is increasing with Δt: at $\Delta t = 10$ minutes, while $\bar{P}$ only reaches 8.7% in experiment 7, it reaches 14.4% in experiment 11, and even 21.7% in experiment 9. These differences based in the topology of the respective temporal networks, but may have implications on potential dynamics on top. For example, the fact that the contact sequences in experiment 9 reach many more agents in the same time interval as experiment 7, implies that spreading phenomena like disease transmission or rumor spreading may be more fast-paced in the setting of experiment 9. There are several considerations to be made when concluding this, as discussed in 4.1. While the black lines show averages over t, the increase of $\bar{P}$ with Δt varies in time—sometimes this growth is faster than in other moments—which is shown in the grey shaded areas. Note that the variations around the black curve are much larger in experiments 9 and 11 than in experiment 7, also seen in panels (a)–(c).

Relating C to existing network metrics. The bottom panels (g)–(i) in Fig. 3 show $C(i, t, \Delta t)$, which is the effect of the removal of agent i on the black curves shown in panels (d)–(f). The curves of all agents i are shown in gray, but we focus on four aspects: the agent with the highest C for $\Delta t = 300$ in red, the same for $\Delta t = 600$ in blue, the same for $\Delta t = 90\%$ of the total time in green, and the agent with the highest degree in the aggregated contact network in the interval $(t, t + \Delta t)$ in black dashed lines. Note that we scanned C based on a few chosen Δt values to highlight particular agents (in red, blue and green), but then plotted these agents over the whole spectrum of Δt values (on the horizontal axis)—'their' Δt value is merely to label them and does not play any further role. We henceforth refer to these agents as the 'red', 'blue' and 'green' agents. The reason for plotting the highest-degree agent (black dashed) is to assess whether having a high degree in the aggregated contact network is be related to the most impactful agents and the various time scales Δt. The answer clearly depends on the experiments: in experiment 7, there is a single curve clearly above all others on all time scales Δt, which means that the red, blue and green agents overlap. With a clear separation to the second and third places, the highest degree agent is not even close to having most impact on larger time scales (note that the vertical scale is logarithmic, which creates a natural convergence of high-impact curves to the right), albeit that this agent remains is in the top 10% of agents with highest C values. In experiment 9, the place for the highest C switches between agents. Up to values of Δt of about 1300 s, a particular agent dominates C (overlapping blue and red), which also happens to be the agent with the highest degree. However, at $\Delta t = 1300$ sec, the green agents takes over the dominance in C, who had a modest impact before that. Such separation reveals the dependency on time scale when assessing the impact of a specific agent on contact sequences. In particular, it confirms the hypothesis that aggregated properties do contain most information for contact sequence centrality at small values of Δt, while they lose their relevance at larger values of Δt. This is highly pronounced in experiment 11. At short time scales, the red curve marks the highest-$\Delta\bar{P}_i$ individual, which also happens to be the highest-degree agent. Around $\Delta t \approx 400$ sec, this individual is surpassed by another (marked in blue), which dominates the spectrum of $\Delta\bar{P}_i$ at most time scales, but is caught up by yet another (in green) at the largest time scale, where the red/highest-degree agent is clearly having a lower impact than many other agents.

Further, Fig. 4a–f show the relationships between $C(i, t, \Delta t)$ at a fixed value of Δt (averaged over t)—denoted by $C(i, t_{av}, \Delta t)$—and six existing network metrics. On the one hand, in panels (a)–(c) we show three such scatterplots: $C(i, t_{av}, \Delta t)$ versus the degree of the agent i in the aggregated network, obtained from the aggregated contact network; the number of (non-individual) events the agent participates in; and the average size of the events (including those including only a single disconnected agent) in the time interval $(t, t + \Delta t)$. Being obtained from the aggregated network, these are all 'static' network properties: they do not provide information on the sequences of interactions. For example, the degree is obtained from the static contact network and frequency or sequences are not taken into account. Likewise, the number of non-individual events an agent has been participating in does not necessarily reflect a wide range of agents—all participated events may be containing only the same people, and the average event size reflects how often the people participates in individual events (i.e., isolated from the rest of the agents). All three metrics (especially the [fully aggregated] degree) correlate relatively well with C for smaller values of Δt, across all three experiments (in colors). Still, there is quite some spread: agents with the highest C are not the ones with the highest network metrics. Also, while experiment 7 has generally smaller degree, the scatter in the C values seems to be equal to the other two experiments. On the other hand, panels (d)–(f) describe three forms of betweenness: event, contact and temporal betweenness (see "Other agent-based topological network metrics" section). A clean relation between C and temporal or event betweenness is barely visible.

Also relevant to the art fair experiment are metadata obtained from questionnaires (for details, see Ref.[42]). We aggregated questions related to the attitudes of the individuals towards the interventions and dangers of the COVID-19 pandemic that was present during the time of the experiment (August 2020), involving questions like "Do you adhere to the 1.5 meter distance rule as issued by the Dutch government?". Participants' reply on a scale from 1 to 7, where a lower score indicates more skepticism on the pandemic and less strict following of the Government imposed COVID-19 rules, while a higher score means the opposite. The average over 22 of such questions is shown at the horizontal axis in panel Fig. 4g. Clearly, there are no significant differences between the experiments, and no relation is visible with C. In panel Fig. 4h, we also plot the age of these participants, also

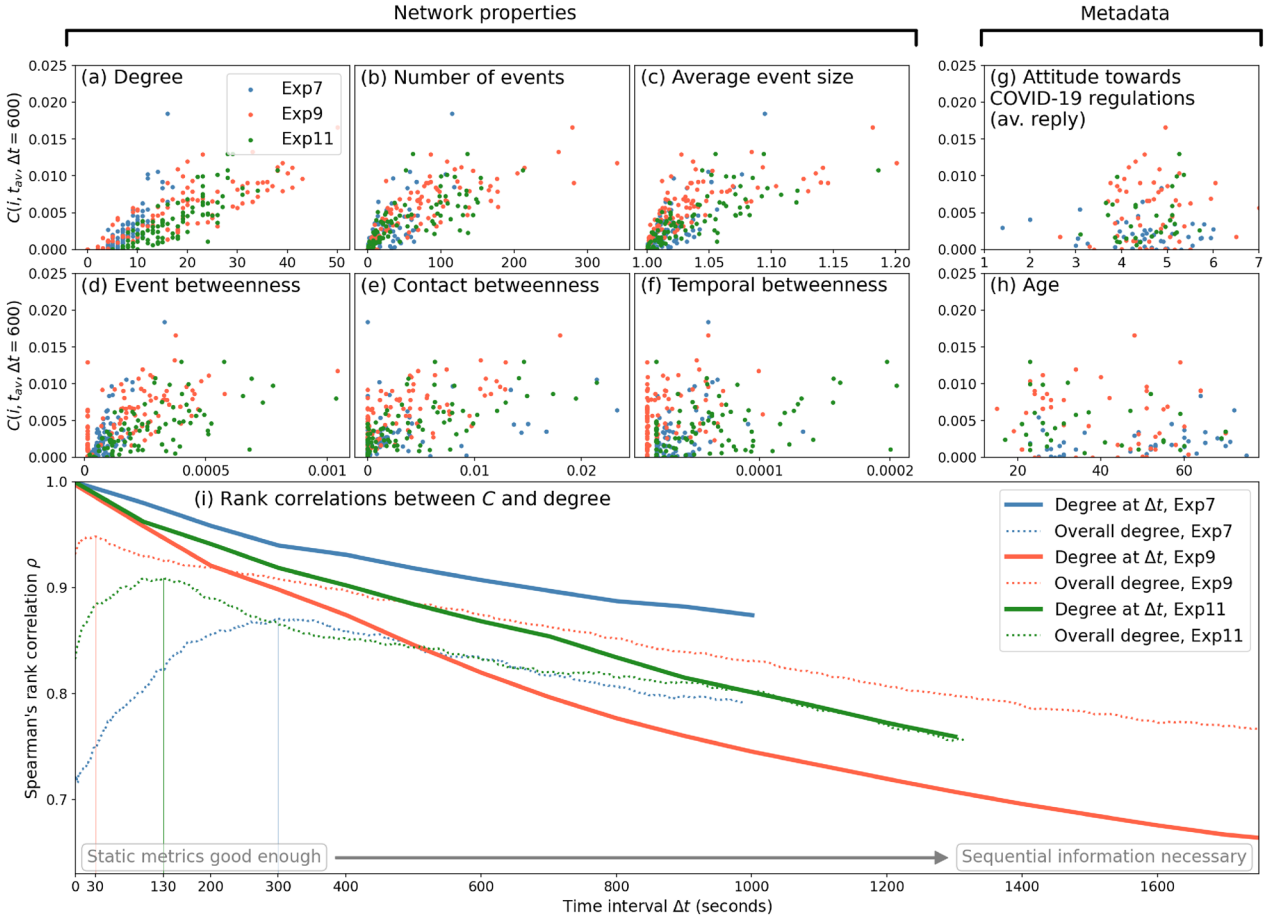

Figure 4. **Panels (a–f)**: Relation between agent's impact on C and various network properties. Different colors denote different experiments. **Panels (g,h)**: Same as in panels (**a–f**), but for two metadata metrics: the inferred attitude of the agent with respect to COVID-19 measures based on questionnaire data, and the agent's age. **Panel (i)**: Spearman's rank correlation coefficients between contact degree and Ci for different values of Δt for each of the three experiments. Thick lines indicate the correlation coefficients when using the average degrees of the agents when looking at interval sizes Δt (as marked by the horizontal axis), while the dotted lines use the overall degree, independent of Δt. Per experiment, Δt values of up to half of the experiment's total time interval are used, to keep enough initialization times t to average over.

yielding no significant relationship with the impact on spreading capability. Not all agents filled in the questionnaires however, resulting in less dots in the metadata panels than in panels (a–f).

Putting together Figs. 3g–i and 4a–f, several relations between network metrics and C can be inferred, but any such relation would be subject to the time scale Δt: for small values of Δt, the static degree may be well suited to describe an agent's impact: having more unique direct contacts means that the agent, potentially, have been in contact with many other agents, which directly affects $\bar{P}$. In turn, therefore, removing that agent would result in a higher C. At higher time scales, however, the effect of sequences becomes more important. Agents with (potentially) smaller initial impact and contact degree may have a larger impact on larger time scales because they serve as conduits of contact for large bubbles or other more subtle temporal structures. This is visible in Fig. 3g–i. To explicitly test the effect of time scale on whether traditional network metrics are sufficient, we show the the Spearman's rank correlation coefficient between the contact degree and C as a function of Δt in Fig. 4i. The overall degree (independent from Δt) is also correlated to C, shown in dotted lines. A clear decay of this correlation with Δt is visible, reflecting that for small Δt, static metrics like the contact degree are sufficient, but for larger time scales, the sequential information becomes important, which motivates a more complex metric like contact sequence centrality C.

Discussion and conclusion

In this paper we have developed and demonstrated a methodology to quantify agent impact on contact sequences in social interactions from the perspective of spreading phenomena. Spreading phenomena are versatile in their nature and types; however, one common element they share is that they all require individual agents as carriers of the entity, and an agent in possession of the concerned entity can pass it on to others that do not have it, leading to the natural expectation that the temporal topology of agent interactions—who interacts with whom, when and in which sequence—will have a profound influence on the spreading dynamics.

In order to quantify this effect, we have proposed a new measure for individual differences in their impact on contact sequences. The measure is obtained by porting all contact information—while preserving the sequences of contacts—into an event graph, and tagging the connections between events by agents' temporal paths. From the event graph, we determine the metric $\bar{P}(t, \Delta t)$, which is the average fraction of agents at time $(t + \Delta t)$ traceable from any agent at time t. Likewise, contact sequence centrality $C(i, t, \Delta t)$ represents an agent i's impact on the contact sequence in terms of how strongly $\bar{P}$ changes when all temporal links belonging to agent i are removed.

For event mapping, we have used the convention that at every (integer) snapshot the agents' contact network within an event is a complete one. For example, for the agents' contact network snapshot at time t, agents 1-and-2 and 2-and-3 may have been recorded to directly interact, but not 1-and-3: in such a case event mapping introduces a direct interaction between agents 1-and-3 that are not originally present in the empirical data. While this is a subtle issue discussed elsewhere in detail[38], we note the value of C is dependent on the choices made in this convention. This is not necessarily a drawback for the concept of contact sequence centrality that we have developed here, but may limits its applicability depending on the time-scale of transmission of a quantity of interest. For example, imagine that the time-scale for transmission is very short relative to the typical duration of an event. Then, in the above example event involving three agents 1, 2 and 3, it will not matter whether 1 and 3 interact directly or not: agent 1 will effectively be in direct contact with agent 3.

Note also that we have calculated both $\bar{P}$ and $\bar{P}_{-i}$ forward in time. This is an obvious choice for spreading dynamics, but for other applications, calculating these quantities backward in time would be more appropriate. For example, when interested in the vulnerability of a specific vendor in a supply chain, the more time-backward dependence the vendor has, the more it will be vulnerable to random failure. In other words, the choice of the direction of time for calculating $\bar{P}$ and $\bar{P}_{-i}$ depends on the application. Note also that the removal of a specific agent in our calculations is merely a conceptual construct to assess its impact on the contact sequence—the actual removal of this agent may have unforeseen effects on the topology of the temporal network that cannot be captured in a conceptual framework.

Contact sequence centrality is a purely topological property, reflecting properties of connections. In other words, agents with high levels of C can be treated as what are colloquially referred to as (behavioral) *superspreaders*[25,27,28], being more central to the contact sequences. Our work serves to highlight the fact that superspreaders are not necessarily those individuals that have most contacts: as shown in Fig. 3g–i, we see that the highest-degree agent starts losing its dominance in C when looking at larger time scales. In particular, increasing the time interval Δt makes the degree less and less determinant for the C values, as shown in Fig. 4i. What sets our metric apart from aggregated (and static) and various temporal metrics is that the exact sequence of the interactions matter: in the example of Fig. 1c, removing agent 3 removes an edge early on in the time series, which has cumulative effects later on. Moreover, reversing the direction of time would yield the same values of all static metrics like degree, and also for temporal betweenness for the aggregated networks, even though it would fundamentally change the values C. Note that our approach had a specific focus on identifying an agent's spreading potential, while many other related approaches exist, each with own scope and (dis)advantages[39]. For example, Yu et al. (2020) combines network embedding with machine learning to assess a node's importance in terms of spreading as evaluated by a SIR-model, which contrasts with our approach relying only on topology[43].

Finally, it is important to realize that topology of the contact sequences alone is not enough for the behavior-epidemiology interface, or for that matter for any dynamical process playing out on top of the temporal network; the process itself may have its own inherent time-scale parameters. Let us elaborate a bit on this point using an example like pathogen spreading, which comes with its own time-scales, such as incubation period, or the time required for an infected agent can itself be infectious. On a given school day, all students in a school may be sequentially traceable to each other as measured by our metrics: $P_i = 1$ for all students i, and $\Delta t = 1$ day. However, the disease may only spread to students directly linked to a patient zero i_0, simply because secondary contacts do not matter because of the incubation time of COVID-19: the direct contacts were not infectious yet. Given this, it is natural to expect that the eventual dynamics of pathogen spreading will be a combination all time-scales involved, including event frequency in the temporal network.

Indeed, to make the current approach relevant for epidemiology, the observation period of contacts among the agents needs to be sufficiently longer than the incubation period, and the time-lag between being infected and being infectious, bringing us to an issue of practicality. While the event mapping is a relatively quick process to compute, depending on the amount of time steps T, the calculation of $\bar{P}$ scales with an extra factor N because the contact sequence tracking [e.g., as in Fig. 1b] has to be done for every agent. However, when computing the C for every agent i, yet another factor N is brought in, as all calculations have to be done fully over when removing i. Hence, the computational complexity of C is of the order $\mathcal{O}(TN^2)$. This means, given that the contact networks of the agents are sampled at 1-s interval, that for making the current method applicable for COVID-19 epidemiology (with an incubation time of about 5.2 days[44] and infectiousness starting time of 12.3 days[45]), further course-graining methods to preprocess the network in the the temporal domain will need to be developed.

In the current paper, we have analyzed an application inspired by the COVID-19 pandemic, as the structure and dynamics of contact networks have proved pivotal to forge a conceptual link between human behavior and virus spread. However, our understanding of human behavior is still limited by a lack of adequate measurement techniques and modeling frameworks. The work presented in this paper showcases an extension of approaches to data analysis, in which experimental manipulations can be tracked to a level of detail that was not previously available. As is the case for any data analytic method, the approach requires existing high-resolution data. Currently, this limits the possibility of applying these calculations to country-wide scales, typically associated with spreading processes such as epidemics. The contribution of this work primarily lies in the characterization of local environments such as supermarkets and museums—it reveals system properties rather than laying out the full spreading infrastructure.

On the basis of the growing number of experimental studies on the topic, this system-characterizing methodology may become available in many situations, which may allow for policy making through the anticipation of spreading vulnerability in future situations. Potentially, the methodology is also used in simulation software that one could use to run through different possible floor plans of an event such as the art fair described in this paper. Policymakers would be in a position to assess different scenarios while planning events. While such techniques are not yet available, we hope that detailed insights into human behavior as obtained through the current analysis may contribute to their development.

In our view, further development of technology and methodology to obtain and analyze direct measures of behavioral contact networks is essential to advance our understanding of human behavior, and to improve the resilience of society in dealing with pandemics.

Data availability

The code for calculating the contact sequence centrality will be made available upon publication. Details about the data of the art fair experiments can be found in Ref.[42].

Received: 6 July 2021; Accepted: 10 February 2022
Published online: 03 March 2022

References

1. Choi, D., Chun, S., Oh, H., Han, J. & Kwon, T. T. Rumor propagation is amplified by echo chambers in social media. *Sci. Rep.* **10**(1), 310 (2020).
2. Stehlé, J. *et al.* Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC Med.* **9**(1), 87 (2011).
3. Kotsakos, D., Sakkos, P., Katakis, I. & Gunopulos, D. Language agnostic meme-filtering for hashtag-based social network analysis. *Soc. Netw. Anal. Min.* **5**(1), 28 (2015).
4. Génois, M. & Barrat, A. Can co-location be used as a proxy for face-to-face contacts?. *EPJ Data Sci.* **7**(1), 11 (2018).
5. Cattuto, C. *et al.* Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLOS ONE* **5**(7), 1–9. https://doi.org/10.1371/journal.pone.0011596 (2010).
6. Stehlé, J. *et al.* High-resolution measurements of face-to-face contact patterns in a primary school. *PLOS ONE* **6**(8), 1–13. https://doi.org/10.1371/journal.pone.0023176 (2011).
7. Masuda, N. & Holme, P. Predicting and controlling infectious disease epidemics using temporal networks. *F1000Prime Rep.*https://doi.org/10.12703/p5-6 *(2013).*
8. Zhang, J. *et al.* The impact of relaxing interventions on human contact patterns and SARS-CoV-2 transmission in China. *Sci. Adv.*https://doi.org/10.1126/sciadv.abe2584 *(2021).*
9. Liu, Q.-H. *et al.* Measurability of the epidemic reproduction number in data-driven contact networks. *Proc. Natl. Acad. Sci.* **115**, 12680–12685. https://doi.org/10.1073/pnas.1811115115 (2018).
10. Dekker, M. M. & Panja, D. Cascading dominates large-scale disruptions in transport over complex networks. *PLOS ONE* **16**, 1–17. https://doi.org/10.1371/journal.pone.0246077 (2021).
11. Blanken, T. *et al. Smart Distance Lab: A New Methodology for Assessing Social Distancing Interventions* (OSF Preprints, 2020). https://doi.org/10.31219/osf.io/mjg2f.
12. Li, A. *et al.* Evolution of cooperation on temporal networks. *Nat. Commun.* **11**(1), 2259 (2020).
13. Isella, L. *et al.* What's in a crowd? Analysis of face-to-face behavioral networks. *J. Theor. Biol.* **271**(1), 166–180. https://doi.org/10.1016/j.jtbi.2010.11.033 (2011).
14. Riolo, C., Koopman, J. & Chick, S. Methods and measures for the description of epidemiologic contact networks. *J. Urban Health Bull. New York Acad. Med.* **78**(3), 446–457. https://doi.org/10.1093/jurban/78.3.446 (2001).
15. Li, A., Cornelius, S. P., Liu, Y.-Y., Wang, L. & Barabási, A.-L. The fundamental advantages of temporal networks. *Science* **358**(6366), 1042–1046. https://doi.org/10.1126/science.aai7488 (2017).
16. Mucha, P. J., Richardson, T., Macon, K., Porter, M. A. & Onnela, J.-P. Community structure in time-dependent, multiscale, and multiplex networks. *Science* **328**(5980), 876–878. https://doi.org/10.1126/science.1184819 (2010).
17. Centola, D. The spread of behavior in an online social network experiment. *Science* **329**(5996), 1194–1197. https://doi.org/10.1126/science.1185231 (2010).
18. Schläpfer, M. *et al.* The scaling of human interactions with city size. *J. R. Soc.* **11**(98), 20130789. https://doi.org/10.1098/rsif.2013.0789 (2014).
19. Kiti, M. C., Melegaro, A., Cattuto, C. & Nokes, D. J. Study design and protocol for investigating social network patterns in rural and urban schools and households in a coastal setting in Kenya using wearable proximity sensors. *Wellcome Open Res.* **4**(84), 17 (2019).
20. Perra, N., Gonçalves, B., Pastor-Satorras, R. & Vespignani, A. Activity driven modeling of time varying networks. *Sci. Rep.* **2**(1), 469 (2012).
21. Scholtes, I. *et al.* Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks. *Nat. Commun.* **5**(1), 5024 (2014).
22. Peixoto, T. P. & Rosvall, M. Modelling sequences and temporal networks with dynamic community structures. *Nat. Commun.* **8**(1), 582 (2017).
23. Dezső, Z. & Barabási, A.-L. Halting viruses in scale-free networks. *Phys. Rev. E* **65**(5), 055103. https://doi.org/10.1103/PhysRevE.65.055103 (2002).
24. Pastor-Satorras, R. & Vespignani, A. Immunization of complex networks. *Phys. Rev. E* **65**(3), 036104. https://doi.org/10.1103/PhysRevE.65.036104 (2002).
25. Kitsak, M. *et al.* Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893. https://doi.org/10.1038/nphys1746 (2010).
26. Wang, S. & Zhao, J. Multi-attribute integrated measurement of node importance in complex networks. *Chaos Interdiscip. J. Nonlinear Sci.*https://doi.org/10.1063/1.4935285 *(2015).*
27. Ahajjam, S. & Badir, H. Identification of influential spreaders in complex networks using HybridRank algorithm. *Sci. Rep.* **8**(1), 11932 (2018).
28. Zhang, D., Wang, Y. & Zhang, Z. Identifying and quantifying potential super-spreaders in social networks. *Sci. Rep.* **9**(1), 14811 (2019).
29. Grabowicz, P. A., Aiello, L. M. & Menczer, F. Can co-location be used as a proxy for face-to-face contacts?. *EPJ Data Sci.* **3**(1), 27 (2014).
30. Kobayashi, T., Takaguchi, T. & Barrat, A. The structured backbone of temporal social ties. *Nat. Commun.* **10**(1), 220 (2019).

31. Mellor, A. The temporal event graph. *J. Compl. Netw.* **6**(4), 639–659. https://doi.org/10.1093/comnet/cnx048 (2017).
32. Mellor, A. EVENT graphs: Advances and applications of second-order time-unfolded temporal network models. *Adv. Compl. Syst.* **22**(3), 1950006. https://doi.org/10.1142/S0219525919500061 (2019).
33. Peel, L., Larremore, D. B. & Clauset, A. The ground truth about metadata and community detection in networks. *Sci. Adv.*https://doi.org/10.1126/sciadv.1602548 *(2017)*.
34. Badie-Modiri, A., Karsai, M. & Kivelä, M. Efficient limited-time reachability estimation in temporal networks. *Phys. Rev. E* **101**(5), 052303. https://doi.org/10.1103/PhysRevE.101.052303 (2020).
35. Torricelli, M., Karsai, M. & Gauvin, L. weg2vec: Event embedding for temporal networks. *Sci. Rep.* **10**, 1–11 (2020).
36. Saramäki, J., Kivelä, M. & Karsai, M. Weighted temporal event graphs. arXiv:1912.03904 (2019).
37. Kivelä, M., Cambe, J., Saramäki, J. & Karsai, M. Mapping temporal-network percolation to weighted, static event graphs. *Sci. Rep.* **8**(1), 1–9 (2018).
38. Dekker, M. M., Schram, R., Ou, J. & Panja, D. arXiv:2107.01651, 1–14 (2021).
39. Hafiene, N., Karoui, W. & Romdhane, L. B. Influential nodes detection in dynamic social networks: A Survey. *Exp. Syst. Appl.* **159**, 113642 (2020).
40. Masuda, N. & Lambiotte, R. *A Guide To Temporal Networks* Vol. 6 (World Scientific, 2020).
41. Kim, H. & Anderson, R. Temporal node centrality in complex networks. *Phys. Rev. E* **85**(2), 026107. https://doi.org/10.1103/PhysRevE.85.026107 (2012).
42. Tanis, C. C. *et al.* Smart Distance Lab's art fair, experimental data on social distancing during the COVID-19 pandemic. *Sci. Data* **8**(1), 179 (2021).
43. Yu, E.-Y., Fu, Y., Chen, X., Xie, M. & Chen, D.-B. Identifying critical nodes in temporal networks by network embedding. *Sci. Rep.* **10**(1), 1–18 (2020).
44. Li, Q. *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England J. Med.* **382**(13), 1199–1207 (2020) (**PMID: 31995857**).
45. He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**(5), 672–675 (2020).

Acknowledgements

Figure 2 is reprinted from[42], for which we acknowledge the authors and thank Matthijs Immink (https://matthijsimmink.com) for the photo of the art fair.

Author contributions

All authors contributed to the research conceptualization. M.M.D. and D.P. conceived the mathematical principles, with help of J.O. M.M.D. performed the data analysis and application to the art fair, with help of T.B, F.D. and D.B. M.M.D. wrote first draft of the manuscript. All authors reviewed the final text.

Funding

This work is part of the project 'Evidence-based effective monitoring and control of COVID-19 after the initial outbreak', which has been funded by the Netherlands Organization for Health Research and Development (ZonMw) with Project Number 10430022010001. This work is also part of the research programme `Improving the resilience of railway systems' with project number 439.16.111. MMD was supported by this research project, which is financed by the Dutch Research Council (NWO) and co-financed by Nederlanse Spoorwegen (NS) and ProRail.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-07384-0.

Correspondence and requests for materials should be addressed to M.M.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022