

The dynamical formation of ephemeral groups on networks and their effects on epidemics spreading

Marco Cremonini^{1,*} and Samira Maghool^{2,*}

¹University of Milan, Department of Political and Social Sciences, Milan, Italy

²University of Milan, Department of Computer Science, Milan, Italy

*marco.cremonini@unimi.it

*samira.maghool@unimi.it

ABSTRACT

In network models of propagation processes, the individual, microscopic level perspective is the norm, with aggregations studied as possible outcomes. On the contrary, we adopted a mesoscale perspective with groups as the core element and in this sense we present a novel agent-group dynamic model of propagation in networks. In particular, we focus on ephemeral groups that dynamically form, create new links, and dissolve. The experiments simulated 160 model configurations and produced results describing cases of consecutive and non-consecutive dynamic grouping, bounded or unbounded in the number of repetitions. Results revealed the existence of complex dynamics and multiple behaviors. An efficiency metric is introduced to compare the different cases. A Null Model analysis disclosed a pattern in the difference between the group and random models, varying with the size of groups. Our findings indicate that a mesoscopic construct like the ephemeral group, based on assumptions about social behavior and absent any microscopic level change, could produce and describe complex propagation dynamics. A conclusion is that agent-group dynamic models may represent a powerful approach for modelers and a promising new direction for future research in models of coevolution between propagation and behavior in society.

Introduction

In this work, we study the effects of the spontaneous formation of ephemeral groups during the declining part of a SIR-type propagation process¹ and the possibility that those temporary groups could ignite a new propagation dynamic. The model takes inspiration from what we learned during the COVID-19 pandemic: increased urban mobility and gatherings have been observed not just as a consequence of the decision of health authorities to remove social restrictions; instead the phenomena has been documented starting right after the peak of the infection and increasing while approaching the lifting of social restrictions². This spontaneous behavior, seemingly anticipating the decisions of health authorities, has effects that make the epidemic dynamics in the declining part a co-evolutionary process with a rich behavioral component still largely unaccounted in research. Most likely reasons for ephemeral groups formation could be the increasing psychological fatigue in complying with rules of social isolation following a period of compulsory seclusion and deprivation of social gatherings and activities³⁻⁵, distrust in health authorities and decay in risk awareness⁶, socio-economic disparities⁷, or seasonality effects like amenable weather conditions favoring gatherings in public places. It is probably safe to assume that a combination of several factors is at play in what appears in many cases a recurrent pattern: generalized strict social distancing has proved to be unattainable during the declining epidemic phase, even in presence of unsafe contagion rates. More generally, a large body of scientific literature exists on collective behaviors that epidemic models still do not fully address^{8,9} and the case of the COVID-19 pandemic makes no exception^{10,11}.

Several recent COVID-19 works have shifted their focus on the conditions characterizing the declining part of an epidemic process, when the rate of contagions sensibly drops and social restriction measures are lifted, and on the combination of causes allowing for a relapse of the epidemic dynamics^{12,13}. However, the vastness of the interplay between epidemic dynamics and behavioral responses remains still largely unexplored. Similarly, in the more general domain of co-evolutionary models of propagation and behavior, the study of dynamic network adaptations and the sheer variety of possible temporal phenomenon is still limited to few cases. This work moves in the direction of integrating an instance of collective behavior into agent-based network models of epidemics, with the specific focus on *ephemeral groups*, as we have dubbed the particular mesoscale structure of our interest. Two research goals have especially driven the work: first, to demonstrate the feasibility of our dynamic grouping technique in simulating network effects able to capture the behaviors produced by ephemeral groups of different characteristics. Secondly, to study and compare the effects of large and small ephemeral groups on the dynamics. In particular, we compare effects produced by scenarios with many small groups, which with the decreasing of their size are less likely to include an infected agent, with scenarios with few large groups.

Results show that a rich variety of dynamics could be reproduced with the combination of two network components, one static representing the traditional aggregated propagation based on slow social network changes and one dynamic representing the dynamic formation of ephemeral groups. Results also suggest that effects of a large group could be equally produced by simultaneous small groups, even in presence of a high rate of non-spreading groups. In the following sections, we first introduce and motivate the notion of ephemeral groups, then we describe the methodology and relevant simulation results for the different model configurations, each representing a stylized case study. We conclude discussing our findings and with some final remarks. This manuscript is completed by *Supplementary Information* presenting details about network characteristics, model states, and algorithms. In addition and of particular relevance is the *Null Model analysis* evaluating our ephemeral groups model with respect to two random models and a special case considering a non-overlapping version of our model.

Ephemeral Groups Formation

The adoption of dynamic networks in epidemic models has frequently focused the attention of researchers on two main aspects: first, how network metrics developed for static networks should be modified when a dynamic network is considered¹⁴; second, the characteristic burstiness of human contact networks, which has been thoroughly investigated in different domains, from human mobility and communication, to physical contacts as the contagion vector^{15–18}. Group formation, instead, has been extensively studied in network science, mostly in the form of generative models and community detection algorithms^{19–21}. However, with respect to the study of microscopic structural features (agent specific), mesoscopic structural features (group specific) influencing a propagation dynamic have generally received less attention^{22–24}. However, some notable works with specific focus on mesoscopic features have recently appeared: in COVID-19 research²⁵, for the most general case of high-order interactions in complex systems²⁶, and regarding high-resolution models of human mixing patterns and mobility, which define contact patterns and adopt mesoscopic descriptions^{27,28}. High-resolution models, in particular, being data-driven, could be the ideal complement of a theoretical work like ours, providing the much needed case studies and real examples of mesoscopic structures necessary for realistic simulations and further model refinements and extensions.

In our model, *ephemeral groups* are mesoscopic structural components defined as spontaneous, temporary, and unstructured. They are mostly composed by unrelated individuals casually gathering together with no central planning, coordination, or organizational rules. They are typically small-to-medium sized and may represent, for instance, occasional parties or other unplanned assemblage in public or private spaces aimed at gregarious activities or random people finding themselves in close proximity, for a limited time span, near points-of-interests, in public transportation, or commercial locations. Ephemeral groups share the properties of possibly form simultaneously, repeatedly, and temporarily mixing individuals, often at random. Different is for large gatherings, which are likely to happen for special social events or circumstances, like conventions, sport competitions, fairs, or seasonal touristic influx in traditional locations. Usually, they are unique events, at least in a certain place and within a time period. Several cases of large social events have been studied with respect to their possible impact on contagions as superspreader events^{29–31}. Ephemeral groups, individually, are not superspreader events, whereas it is possible for their combined effect. In addition, they are more difficult to study empirically than single large events. The relatively few exceptions have mostly regarded the case of people in public transportation, with data collected from turnstiles^{32–35}, and scientific experiments involving limited groups of participants equipped with tracking devices (e.g., students on campus, visitors of museum exhibitions)^{36,37}. We are interested to study the possible effects that those ephemeral groups may have on a possible new surge in a propagation process. The model has two components, which co-evolve producing mutually dependent network effects on the dynamic:

- *Microscopic static features*: An extension of the traditional Susceptible-Infected-Recovered (*SIR*) model based on a static proximity network and simple contagion mechanism. This part has been originally presented in Ref.³⁸ to study the epidemic dynamics under the assumption of different rates of Mild/Acute infected. It provides the base epidemic model and static contact network. Details about the basic epidemic model and model epidemic states are provided in Supplementary Information, Section S1 and S2.
- *Mesoscopic dynamic features*: The dynamic formation of ephemeral groups is modeled as a choice of randomly selected agents that, starting from a given time point in the declining epidemic phase (more on the criteria for selecting when to start ephemeral groups in Supplementary Information, Section S2, subsection *Ephemeral groups creation pseudocode*), dynamically link pairwise in groups having the following characteristics: a variable rate of agents is involved in groups; groups are randomly created and possibly repeated for more time steps, each group may or may not include infectious agents (i.e., the probability to include spreading agents is increasing with the group size).

Fig. 1 shows a schematic representation of ephemeral groups effects studied in this work. On the timeline of simulation steps (time steps), the network dynamic is described with the progression of three characteristic steps:

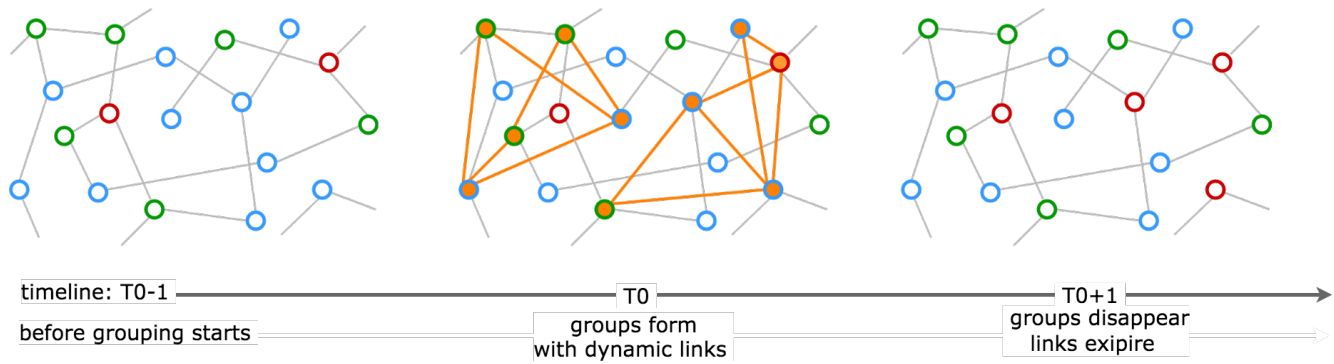


Figure 1. The contact network is described through static links and agent states. *Red* is for Infected (e.g., spreading agents, infected individuals in disease epidemics, adopters in product adoption, spreaders in opinion diffusion), *green* is for Recovered (e.g., agents no longer spreading, immunized, stifter, contrarian, antagonist, deluded are possible definitions in different contexts), and *blue* is for Susceptible (e.g., agents possibly becoming spreading actors, ignorant, undecided, laggard); filled in *orange* are agents belonging to an ephemeral group.

- **Before ephemeral groups formation** (time step $T0 - 1$): The system configuration is the product of the static contact network evolution and agents' state transitions. Specifically, by focusing on the declining part of the dynamic, we consider the system as slowly evolving, with few agents still changing state. In the example of Fig. 1, the two red agents (Infected, for simplicity) are no longer spreading because connected only to green agents (Recovered). In the bottom right corner, a blue agent (Susceptible) appears as disconnected from the visible portion of network.
- **During ephemeral groups dynamic linking** (time step $T0$): Two groups of five agents with two new links for each agent are showed marked in orange in the central vignette. Node selection for each group is random, in this version of the model. Dynamic links are assumed to be intragroup and randomized as well. Propagation from red (I) to blue (S) agents is, therefore, a first order effect of grouping only between agents belonging to the same group. Propagation to external agents is a second order effect. For simplicity, no social learning or herding effects on propagation have been assumed. Ephemeral groups are assumed to have a typical single time step lifetime. Different configurations based on the repetition of the dynamic grouping for more time steps have been simulated and studied.
- **After ephemeral groups dissolve and dynamic links expire** (time step $T0 + 1$): The creation of new links inside ephemeral groups logically reshuffles the topology of the contact network, leading to different effects based on the group composition. In the example of Fig. 1, one group has a red (I) agent among its members, therefore possibly spreading the phenomena, while the other has none, being inactive. Spreading could be enhanced by shortening paths, connecting agents previously disconnected, or offering new contacts to isolated spreading agents, as the third vignette shows with the red agent, previously isolated between two green ones, been able to turn red the disconnected bottom right blue agent and another blue agent, from which the propagation could start over on the static network.

Materials and Methods

Model

We present a stochastic group-based network model focusing on the declining part of a schematic SIR-type propagation process in a society and considering the formation of ephemeral groups creating new dynamic links. The model is an extension and a generalization of a previously presented model³⁸. It extends the previous model by studying the conditions of ephemeral groups leading to a second epidemic dynamic under different scenarios. It is a generalization because the analysis and conclusions have a broader aim than previous COVID-19 research, discussing cases of coevolution between propagation and behavior, when the behavioral component is expressed by a combination of individual (microscale model) and dynamic groups (mesoscale model). We have also simplified the analysis of agents' state transition dynamics in order to better focus on the role of ephemeral groups from the previous extended SIR model (more details in Supplementary Information, Section S2). This way, combined with the assumption of fixed population, we have also restricted the effects produced by ephemeral groups only to a second propagation dynamic. Removing one of these assumptions and allowing for a renewal of the sub-population of agents that could turn into spreading ones (i.e., susceptible turning infected) would make possible to sustain a series of propagation dynamics through the formation of ephemeral groups. For this study, we adopt a constant population of agents $N = 10000$ connected on

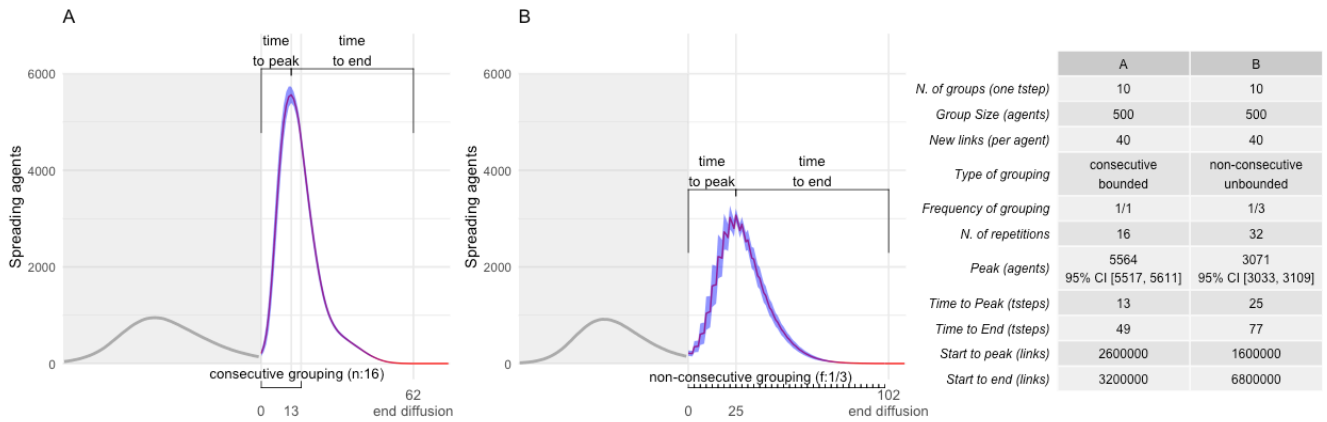


Figure 2. In panels A and B, characteristic features used in reporting the results are shown. The *peak* of the process represents the maximum number of Infected. The peak time step is indicated on the x axis with the offset from the origin (+13 in panel A and +25 in panel B). *Time to peak* and *time to end* represent duration (in time steps). Asymmetry of the curve and dependence on different factors makes it useful to measure the two parts independently. *Start of grouping* and *end of diffusion* are typical points in the timeline, respectively, indicated with 0, corresponding to the flex of the curve and the first time step of dynamic grouping, and a point that varies for each configuration, by convention assumed as the time step when the majority of trials has no Infected agent left (indicated as an offset from the flex, +62 in panel A and +102 in panel B). *Grouping pattern* and *duration* are indicated with horizontal brackets. For panel A, grouping has been repeated for 16 consecutive time steps; for panel B, grouping has a 1/3 frequency and run from start to end of diffusion. In both cases, we see that the peak of the dynamics, while depending on grouping, may occur before grouping ends. The table lists the parameters that describe each configuration. We will use the notation $\#G$ for the *number of groups* and $(y_1 = \text{group size}; y_2 = \text{new links}; y_3 = \text{grouping pattern})$ to indicate a configuration, where the grouping pattern could be a finite number of steps n or a frequency f . The frequency could indicate consecutive grouping $f = 1/1$ or periodic grouping $f = 1/m$ (i.e., $f = 1/3$ is the value we used in tests). For example: given $\#G=10$, $(500; 40; 16)$ identifies the configuration of panel A, $(500; 40; 1/3)$ identifies the configuration of panel B.

an abstract network with scale-free characteristics (Supplementary Information, Section S1 provides specific information about the network generation algorithm, the degree distribution, and some typical network centrality measures). The results of 160 different model configurations are presented, all referring to cases producing a recognizable response on the overall dynamic. They have been run through simulations and for each one we took a sample of 50 valid trials. A trial is considered *invalid* when the second wave was absent or insufficient to represent an epidemic dynamic (i.e., we assumed as the exclusion criteria to have less than 100 additional infected agents at the peak).

Fig. 2 shows two panels, A and B, representing the general classes of propagation dynamics studied in this work, and a table containing the characteristic parameters analyzed for all the configurations and presented in the results. The main difference between panel A and panel B lays in the pattern of grouping repetition: the *grouping frequency* (i.e., the number of grouping time steps over a certain interval) and the *total number of grouping time steps*. For the grouping frequency we have considered two cases: *continuous* when grouping occurs at consecutive time steps and *periodic* when grouping occurs at non-consecutive time steps. For the number of grouping time steps, we have considered two cases as well: *bounded* when a fixed number of grouping time steps is set and *unbounded* when grouping is allowed until the exhaustion of the propagation dynamics. In Fig. 2, panel A represents a continuous and bounded propagation dynamic, in this case 16 are the consecutive grouping time steps, indicated with $n = 16$. Instead, panel B shows a periodic and unbounded propagation dynamic, with grouping occurring at a rate of one every three time steps, indicated as $f = 1/3$, which produces the typical sawtooth shape of dynamic processes characterized by cyclic reinforcements followed by decline phases.

Panels A and B introduce several elements of the dynamics we have studied. The first part of the propagation dynamic, marked with the grey area, is useful only to have a general overview of the full process and of the typical proportions between the two phases of the propagation dynamics (several other cases with a second dynamics smaller than the first have been simulated, but in this work we focus specifically on those producing a large response). Other than that, the first part of the dynamic is not particularly relevant, representing equal initial conditions for ephemeral groups among all cases. Average number, with 95% CI, of Infected, Susceptible and Recovered agents at the start of ephemeral grouping are in Table S2 of Supplementary Information. In following figures, the first epidemic wave will be omitted. A detail to highlight is that we studied the second propagation dynamic as the result of a continuous propagation process that produces a first dynamic, then,

when a threshold on the number of spreading agents (I) is reached in the declining phase, it triggers the dynamic grouping mechanism and a flex is produced, starting the second dynamic. To be more specific, for each configuration, we have aligned all trials with respect to the flex point in order to normalize the time-dependent variability and then averaged over the trials. The first dynamic (i.e., the grayed area up to the flex of panels A and B) is independent from the grouping mechanism and shared by all configurations, with a variability within narrow stochastic bounds.

The cases we analyzed have been categorized in scenarios, each defined by the number of ephemeral groups produced per time step ($\#G$), and for each scenario in configurations. Together, group size, number of new links per agent, the already introduced grouping pattern, and the number of groups uniquely identify each configuration presented in the results. More specifically,

- **group size** represents the number of agents belonging to a group. In this study, we assume that every configuration has groups of equal size. When grouping is repeated for more time steps, at each time step groups are recreated randomly. Overlaps between groups are possible and in general high-order coupling could have relevant effects on the dynamics. For this reason, we tested a non-overlapping variant. The model analysis of statistical significance is presented in Supplementary Information, Section S4, showing that, in our case, group overlaps seems to have a negligible effect.
- **new links per agent** represents the number of new links each agent randomly creates. Dynamic links are only intragroup and have a lifetime of one time step. When grouping is repeated, continuously or periodically, dynamic links are recreated every grouping time step. In this study, we assume same number of new links per agent for each configuration. Dynamic links are removed when a group dissolves.
- **number of groups** represents how diffuse the ephemeral grouping behavior is assumed to be in the population. As a rule of thumb, we assume an inverse proportion between number of groups and their size, that is, we investigated cases of many small groups and of few large groups in order to compare the behavior. The total number of agents involved in ephemeral groups is then calculated as $N. \text{ of groups} \cdot \text{group size}$, while the total number of dynamic links for each grouping time step is $N. \text{ of groups} \cdot \text{group size} \cdot \text{new links per agent}$.

Pseudocode describing the algorithms could be found in Supplementary Information, Section S2.

Results

Model configurations have been divided in four scenarios, each one assuming a different number of groups per time step, $\#G = (1, 10, 100, 1000)$. The choice of these values has been driven by the extremes: on the one side we wish to test the case of a single large group ($\#G = 1$), which mimics a large gathering of loosely connected agents, and on the other the case of a multitude of small groups ($\#G = 1000$), tightly connected by dynamic links but in large majority inactive because without spreading agents. The other two values ($\#G = (10, 100)$) represent intermediate scenarios. For each scenario, eight grouping patterns have been considered, six bounded with number of time steps $n = (1, 2, 4, 8, 12, 16)$ and two unbounded with frequency $f = (1/3, 1/1)$.

Base epidemic and ephemeral groups simulation settings are listed in Table S2 of Supplementary Information. For convenience, we report here the initial conditions of agent states when ephemeral groups start. Values are calculated from the trials taken from a sample of configurations and are in the form *average [CI 95%]: Infected: 140 [142.25,137.75]; Susceptible: 7087 [7122.24,7051.76]; Recovered: 2773 [2806.25,2739.85]*. For each scenario, eight grouping patterns have been considered, six bounded with number of time steps $n = (1, 2, 4, 8, 12, 16)$ and two unbounded with frequency $f = (1/3, 1/1)$.

Consecutive grouping, bounded and unbounded

We first look at results from configurations with consecutive grouping time steps, bounded in the number of repetitions of group formation with $n = (1, 2, 4, 8, 12, 16)$ and unbounded (frequency $f = 1/1$). Configurations with non-consecutive grouping (frequency $f = 1/3$) have peculiar characteristics we discuss later. Results are first presented in Fig. 3 to give a broad overview of the outcome from all configurations and to present observations based on peak values mostly. Then, a more detailed analysis focusing on a subset of configurations and introducing productivity based and time based metrics is summarized in Table 1.

Peak analysis

By considering peak values of Infected agents reached by the different cases, we compare the performances in different scenarios, grouping patterns, and number of dynamic links created up to the peak value.

Very large peak values, in our experiments close or above 6000, mean that the propagation process reached the maximum extension and was capped by the population size and the proportion of agents that cannot be turned into spreading ones (Recovered). Configurations in all scenarios have easily reached such strong responses in a range of parameter values (group size, number of new links per agent, and number of grouping time steps), either with few loosely connected groups (e.g., lesser

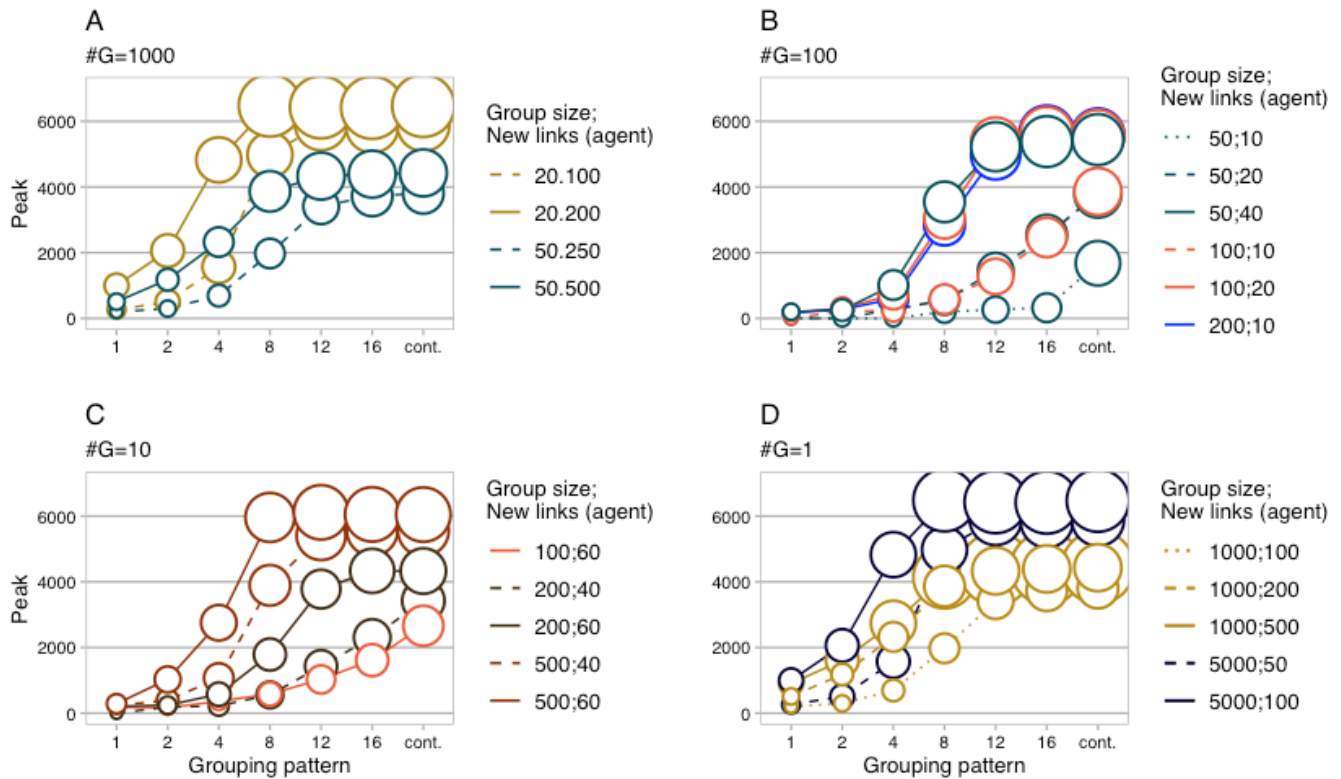


Figure 3. Six model parameters are represented for a total of 140 configurations run through multi-agent simulations (the 20 cases with $f = 1/3$ are not included here). For each configuration, a sample of 50 trials has been taken, aligned with respect to the flex, and averaged. Each circle represents a single configuration (28 in panel A, 42 in panel B, 35 in panel C and D). Parameter description with the corresponding graphic element follows. **Panel:** number of groups created per time step; **x axis:** grouping pattern for six consecutive and bounded cases $n = (1, 2, 4, 8, 12, 16)$ plus the case consecutive and unbounded $f = 1/1$; **y axis:** peak size as number of spreading agents; **color and linetype:** group size and new links per agent; **area of circles:** total new links ($\times 1000$) created by all groups from start of grouping (flex) to the peak (new links possibly created after the peak are not part of the metric). From panels A-D, it derives that configurations that plateaued reached the stationary peak level within 8-12 consecutive grouping repetitions, while grouping patterns of 16 or unlimited time steps are mostly relevant for configurations that produced a weaker response.

than $1/10$ for the ratio of new links per agent over the group size) or with many small tightly connected groups (e.g., ratio of new links per agent over the group size greater than $1/10$).

Therefore, from Fig. 3, a first observation is that a strong propagation dynamic cannot be excluded or it cannot be considered more or less likely only by observing whether the scenario is composed by few large but loosely connected groups (as for panels C and D) or many small but tightly connected groups (as for panels A and B).

Secondly, a plateau is not been reached only as a consequence of bounded network size, but it could be the result of a spontaneous behavior of the propagation process. This is showed in panels C (see configuration 200;60) and particularly in panel D (see the whole series 1000;100, 1000;200, and 1000;500), where peaks reach a medium level around 4000, far from the dominant effect of the limited network size (at that level, the rate of Recovered and Susceptible agents is around 30%). More specifically, in panel D, the series with group size equals to 1000 shows that the peak stays around 4000 for a large range of new links per agent (from 100 to 500). It also exhibits a striking correspondence with the behavior of the series with 5000 as group size reaching the peak value with 8 consecutive grouping time steps and similar number of total new links produced by the group, then plateaued. Other cases in different scenarios behave similarly (e.g, the 50;10 of panel A or the 100;60 of panel C). Another set of cases (i.e., 20;5 of panel A; 100;10, 50;20 and 50;10 of panel B; 200;40 and 100;60 of panel C) have a slower growth and reach the peak value only when we tested with an unbounded number of grouping time steps (for those cases, more than 20 and up to 40 consecutive grouping time steps were necessary to reach the peak).

A positive correlation between the peak level and the total number of links created by groups up to the peak ($\#G \cdot \text{group size} \cdot \text{new links per agent} \cdot \text{time to peak}$) could be observed, but the relation is not linear and could be capped. Series presented in

panel B of Fig. 3 have an almost perfect correlation for all grouping patterns (i.e., 50;40, 100;20, and 200;10 overlaps, as well as 100;10 and 50;20). In panel A and C, the correlation is still clear but, for each configuration, only up to the grouping pattern that plateaued. To this regard, panel D is the most interesting, with the capped peak of the 1000 series. Configuration 200,40 of panel C provides another example of series that plateaued at mid-range of peak values.

Efficiency of dynamic linking

Efficiency is a typical property of a process measured as the ratio between the output of the process to the input, representing which fraction of output could be obtained by consuming a unit of input. Expressed as costs or assets for the input, and number of items produced by the process for the output, an *efficiency metric* is useful for comparing processes and rank them based on how efficiently they use the resources, monetary or operation. Said differently, an efficiency metrics measures the *marginal gain of resources*.

In our model, we have observed that the peak of the epidemic propagation depends on the number of dynamic links up to the peak time step. Therefore, a measure of efficiency of ephemeral grouping processes could be done by considering the total number of Infected agents created during the time to peak phase as the output, and the total number of dynamic links created during the time to peak phase as the input. The model configuration that needs fewer dynamic links to achieve a new Infected agent is the most efficient. Eq. 1 shows the formula.

$$K_{Eff} = \frac{\sum_{i=t_0}^{t_{peak}} \text{Infected}(i) - \sum_{i=t_0} \text{Infected}(i)}{\sum_{i=t_0}^{t_{peak}} \text{dynLinks}(i)} \quad \text{with } \text{dynLinks}(i) = \#G \cdot \text{group size} \cdot \text{new links per agent} \quad (1)$$

More precisely, the K_{Eff} metric is an approximation of the grouping process efficiency, because it does not distinguish between new Infected agents directly produced by dynamic links and those instead created by the epidemic dynamics on the static contact network. However, considering that it is only thanks to the ephemeral groups mechanism that the epidemic has produced a second wave, in first approximation we assume that all new Infected agents are, directly or indirectly, produced by the grouping process.

	N. of groups #G=1000				N. of groups #G=100						
	50;10	50;5	20;10	20;5	200;10	100;20	50;40	100;10	50;20	50;10	
8	1.62	1.91	1.10	0.37	8	1.68	1.82	2.12	0.55	0.52	0.15
12	1.67 ▲	1.95 ▲	1.96	0.63	12	2.02	2.18	2.12	0.95	1.09	0.23
16	1.70	1.77	1.71	1.24	16	1.84 ▲	1.96 ▲	2.02 ▲	1.55	1.50	0.28
1/1	1.68	1.81	1.76 ▲	1.39 ▲	1/1	1.94	1.93	2.04	1.68 ▲	1.65 ▲	0.78 ▲
avg.	1.67	1.86	1.63	0.91	avg.	1.87	1.97	2.07	1.18	1.19	0.36
peak ▲	6816	5999	5284	3604	peak ▲	5652	5635	5444	3861	3783	1678
time to peak	8	13	15	26	time to peak	14	14	13	23	22	40
time to end	46	47	50	69	time to end	47	49	52	62	65	92
links to peak (x1000)	4000	3250	3000	2500	links to peak (x1000)	2800	2800	2600	2300	2200	1950

	N. of groups #G=10					N. of groups #G=1					
	500;60	500;40	200;60	200;40	100;60	8	5000;100	5000;50	1000;500	1000;200	1000;100
8	2.42 ▲	2.34	1.73	0.67	0.94	8	1.59 ▲	2.42	1.00	2.32	2.30
12	2.00	2.20 ▲	2.52	1.34	1.24	12	1.57	2.10 ▲	0.78 ▲	1.93 ▲	2.72
16	1.97	2.09	2.18 ▲	1.69	1.55	16	1.57	2.07	0.78	1.93	2.25 ▲
1/1	1.97	2.08	2.05	1.86 ▲	1.62 ▲	1/1	1.58	2.09	0.78	1.95	2.30
avg.	2.09	2.18	2.12	1.39	1.34	avg.	1.58	2.17	0.84	2.03	2.39
peak ▲	6115	5564	4334	3424	2662	peak ▲	6483	5887	4416	4392	3807
time to peak	10	13	16	22	26	time to peak	8	11	11	11	16
time to end	49	49	54	73	73	time to end	48	48	53	53	56
links to peak (x1000)	3000	2600	1920	1760	1560	links to peak (x1000)	4000	2750	5500	2200	1600

Table 1. Configurations are categorized by number of groups #G in the four quadrants, then by series (*group size*, *new links per agent*) in columns, and finally by grouping patterns, here only $n = (8, 12, 16)$ and $f = 1/1$, in rows. The values of the first four rows are those of the $K_{Eff}(\times 1000)$ efficiency metric, with average efficiency of the ephemeral grouping process in the fifth row. The black triangle (▲) corresponds to the grouping pattern that reached first the peak for each configuration. The last four rows of each quadrant are referred to the pattern with the black triangle ▲ and show: the peak value, the two time based metrics *time to peak* and *time to end*, and *links to peak (x1000)*, representing the number of links (x1000) produced by groups from start of grouping to peak. For example, first model configuration of the first quadrant (50;10 of #G=1000): the ▲ is on the row corresponding to grouping pattern $n = 12$, therefore values of the last four lines should be meant as referred to that pattern.

Table 1 presents the values of $K_{Eff}(\times 1000)$ (in the first five rows of each quadrant), so that different model settings (pattern, configurations, scenarios) could be compared in terms of efficiency of the conversion of dynamic links into Infected agents. The last four rows of each quadrant give information related to the pattern first reaching the peak (peak value, time to peak, time to end, and links to peak).

Values $K_{Eff}(\times 1000) < 1$ represent cases of slow and ineffective propagation, whereas values $K_{Eff}(\times 1000) > 2$ are associated to fast and efficient cases. Next, the positive correlation between the peak value and the number of dynamic links created in the time to peak interval is confirmed for all series with one exception: (1000;500). As a matter of fact, the whole (1000;*) series is a special case, because it shows how the marginal gain in Infected at the peak could decrease with the increasing density of dynamic links: (1000;100) produces the lowest peak but with highest K_{Eff} , then the peak increases for (1000;200) but with a reduced marginal gain, until the (1000;500) case that shows a small marginal gain and a clearly inefficient propagation, up to the point of losing the positive correlation showed by the other cases.

Another correlation that is observed is that *time to peak* and *time to end* are negatively correlated with the peak value and the *links to peak* ($\times 1000$). This is not a new result in general, on the contrary, it is typical of SIR-type dynamics^{1,39}. What is possibly interesting is the fact that the relation between peak value, number of links and duration of the propagation process could be controlled, to some extent, by the grouping mechanism. This offers interesting possibilities, that we further investigate in the next section.

Non-consecutive grouping

Table 2 shows some results from a slightly different set of experiments aimed at investigating the effects of periodic ephemeral groups that activate with a non-consecutive pattern. What we present is our first take, initial results, and analysis of an unexplored group model of propagation dynamics and behavior, to the best of our knowledge, which, we guess, could have numerous interesting variants and applications to many areas. Non-consecutive grouping obviously reduces the total number of dynamic links created in a certain time span with respect to the consecutive case, but we know that the peak is correlated with the number of links created up to the peak, so a first result was to compare the peak analyses in the two cases, consecutive and non-consecutive. Secondly, periodic grouping may produce, as we have showed in Fig. 2 for $f = 1/3$, the typical sawtooth shape of a function that alternates fast growth with a slower decline due to the fast rise of new spreading agents produced by groups followed by the slower dynamic on the static contact network where the rate of conversion from spreading agents (I) to inactive agents (R) is higher than those from potentially spreading (S) to spreading. Therefore, an interesting question regards the possibility to prolong or even fully sustain a propagation dynamic by means of a proper tuning of periodic grouping. Table 2 offers some insights. Comparing peaks analyses:

- with $f = 1/3$, for each group size, the peak value and the number of links up to the peak are still positively correlated, but the correlation is lost among series of different group size;
- considering the difference with consecutive cases, peaks for $f = 1/3$ fall faster than the link numbers, with the exception of cases with $\#G = 1$; this implies a reduced marginal efficacy of new links, confirmed by negative differences of the K_{Eff} index;
- in $\#G = 1$, K_{Eff} differences are all positive, except for (1000;100) and peaks do not fall faster than link numbers.

Instead, with respect to the ability of a non-consecutive grouping pattern to extend the duration of the propagation, results show *time to peak* having increments from 50 to more than 100%. This was expected because it mostly depends from the reduced total number of links from start to peak. The more interesting result is that *time to end* has large increments, from 30 to more than 100%. Notable is that those are increments with respect to the consecutive case $f = 1/1$ that repeats grouping until the end of propagation and creates three times the number of links of $f = 1/3$. Again an important case where more links do not produce more propagation, in fact after the peak $f = 1/1$ has a negative marginal gain with respect to $f = 1/3$.

Discussion

In this work, we have proposed a new dynamic group-based network model of a propagation process, specifically focusing on modeling mesoscopic constructs. Ephemeral groups are the composite subjects of our model. Model simulations, analysis of results and Null Model analysis (presented in Supplementary Information, Section S3) permit us to draw some conclusions.

First, by considering where we started from, an epidemic model inspired by COVID-19 pandemic, we observe that ephemeral groups could be a relevant factor in epidemic dynamics when people return to socialize freely. Large gatherings resulting in superspreading events have been frequently discussed, as well as repetitive gatherings protracted for long periods, such as in schools or workplaces. Ephemeral groups are instead more elusive to a rigorous assessment of the risk they pose. To this regard, future models may shed new light on the relevance of groups on epidemic dynamics. An outcome of our

N. of groups #G=1000					N. of groups #G=100						
1/3	50;10	50;5	20;10	20;5	1/3	200;10	100;20	50;40	100;10	50;20	**50;10
<i>diff %</i>	1.98	1.84	1.39	0.53	<i>diff %</i>	1.15	1.83	2.07	0.84	0.75	0.14
peak	18.82	-1.05	-14.87	-41.55	peak	-38.53	-7.22	-0.19	-28.93	-37.01	-61.05
<i>diff %</i>	5087	3232	2494	694	<i>diff %</i>	2733	2845	2883	820	918	125
time to peak	-25.37	-46.12	-52.80	-80.74	time to peak	-51.65	-49.51	-47.04	-78.76	-75.73	-92.55
<i>diff %</i>	13	28	30	51	<i>diff %</i>	31	28	25	52	48	1
time to end	62.50	115.38	100.00	96.15	time to end	121.43	100.00	92.31	126.09	118.18	-97.50
<i>diff %</i>	59	71	88	136	<i>diff %</i>	77	79	80	113	115	202
links to peak (x1000)	28.26	51.06	76.00	97.10	links to peak (x1000)	63.83	61.22	53.85	82.26	76.92	119.57
<i>diff %</i>	2500	2250	2000	1700	<i>diff %</i>	2200	1800	1600	1500	1600	50
<i>diff %</i>	-37.50	-30.77	-33.33	-32.00	<i>diff %</i>	-21.43	-35.71	-38.46	-34.78	-27.27	-97.44

N. of groups #G=10					N. of groups #G=1						
1/3	500;60	500;40	200;60	200;40	100;60	1/3	5000;100	5000;50	1000;500	1000;200	1000;100
<i>diff %</i>	2.38	2.10	1.83	0.98	0.94	<i>diff %</i>	2.09	2.26	1.15	2.13	2.08
peak	13.80	-3.61	-13.72	-29.61	-29.77	peak	32.49	4.15	37.72	4.80	-13.06
<i>diff %</i>	4074	3071	1931	848	651	<i>diff %</i>	4918	3708	2743	2697	1735
time to peak	-33.38	-44.81	-55.45	-75.23	-75.54	time to peak	-24.14	-37.01	-37.88	-38.59	-54.43
<i>diff %</i>	16	25	31	45	52	<i>diff %</i>	13	18	13	16	28
time to end	60.00	92.31	93.75	104.55	100.00	time to end	62.50	63.64	18.18	45.45	75.00
<i>diff %</i>	72	77	93	117	131	<i>diff %</i>	63	76	113	117	110
links to peak (x1000)	46.94	57.14	72.22	60.27	79.45	links to peak (x1000)	31.25	58.33	113.21	120.75	96.43
<i>diff %</i>	1800	1600	1320	1280	1020	<i>diff %</i>	2500	1750	2500	1200	1000
<i>diff %</i>	-40.00	-38.46	-31.25	-27.27	-34.62	<i>diff %</i>	-37.50	-36.36	-54.55	-45.45	-37.50

** Series 50,10 for $f=1/3$ produces a negligible second dynamic

Table 2. Table layout is similar to Table 1. For each quadrant, the values of the first row are those of the $K_{eff}(\times 1000)$ index for the non-consecutive grouping pattern with $f = 1/3$. The following *diff %* row, show the percent difference with the corresponding values of average efficiency, (*avg.*) row of Table 1. All other *diff %* rows refers to corresponding rows of Table 1.

work is that, at least in theory, for several and diverse configurations, ephemeral groups have the power to restart and sustain an epidemic dynamic when it was headed to expiration, in the absence of contextual changes (e.g., new more transmissible pathogens). Health authorities should exercise prudence in face of the inevitable uncertainty regarding the effects of gatherings and the difficulty of scientists in collecting precise measurements to feed epidemic models.

Secondly, the *time to end* metric seemingly confirms the hypothesis that a periodic dynamic grouping mechanism could deeply influence the propagation process and could sustain it for an extended period. This result could remind of known behavioral and cognitive effects from social learning theory⁴⁰ or awareness models⁴¹, all implying that periodic reinforcements could sustain an otherwise decaying process. A novelty of our work is to have derived some initial quantitative results through simulations of a group model, with the analysis suggesting that balancing the alternation of growth and decay phases, rather than simply the number of reinforcement events, is a key factor and it would be the basis for new and possibly interesting future research in this area.

Another outcome that looks interesting comes from the Null Model analysis. It suggests a change of the relevance of groups through the different scenarios. Many small groups, most of them with no internal epidemic propagation because without infected members, seem to act similarly to a random linking model. But they could be nevertheless effective in sustaining an epidemic dynamic. In real settings, they could possibly join to form larger groups. In larger groups the difference with random linking models emerges increasingly clear. Simulation data show that the ephemeral groups model reaches often a slightly higher peak with fewer active links (i.e, links between an Infected and a Susceptible agent) than corresponding random cases. This seems to indicate that groups seed and sustain a more efficient epidemic propagation. A dynamic of groups aggregating to form larger compounds represents a behavioral phenomena potentially at the base of particularly meaningful propagation models. On the practical side, not just the creation of ephemeral groups and their characteristics are relevant in a propagation phenomena, but the temporal evolution of groups organization is important as well.

Other observations are probably less relevant, but still worth mentioning. From Table 1, we obtained a detailed description of the dynamics produced by the different cases of consecutive grouping. The total number of links created up to the peak is the factor that determines the peak level (positively correlated) and the time needed to reach it (negatively correlated), which implies the existence of feedback loops in the dynamic. An important limit to the correlation between peak and number of links is the case of a propagation process that plateaued, which happens when the ability of dynamic links to produce new

spreading agents drops to values of the $K_{Eff}(\times 1000) < 1$. That could happen due to the limited size of the population (i.e., the trivial case of saturation) or for contextual conditions of the propagation process for which, up from a certain level, increasing densities of dynamic links accentuate inefficiency in propagation and a decreasing marginal gain of links. The correlation is lost if the total number of links produced from the start of grouping to the end of propagation is considered, rather than only those up to the peak. Finally, non-consecutive grouping reveals interesting scenarios. The analysis of results from Table 2 disclosed a complex interplay between the two dynamics running in parallel, the one on the static contact network and the one of dynamic grouping. On the one side, the reduced marginal efficacy of dynamic links of the non-consecutive cases produces less new infected agents, and on the other it changes the balance between the two dynamics, shifting the weight in favor of the rate of agents recovering. This makes peaks falling, with a rate that accelerates with the loss of efficacy of dynamic links.

A concluding remark is that by studying this model, we had the clear feeling of how vast and rich could become models based on dynamic groups. Moving from typical epidemic models to models designed for the propagation of ideas, news, or products will keep some common features, but will introduce new set of assumptions, radically changing the dynamics. High-order coupling, here barely touched in Supplementary Information, group membership persistence, heterogeneous groups, time dependent agent behavior and social reinforcement, the influx and outflux of new agents, as well as the representation of different cultures in agent populations could all become key factors of future challenging as well as interesting models.

References

1. Hethcote, H. W. The mathematics of infectious diseases. *SIAM review* **42**, 599–653 (2000).
2. Alfaro, L., Faia, E., Lamersdorf, N. & Saidi, F. Social interactions in pandemics: fear, altruism, and reciprocity. Tech. Rep., National Bureau of Economic Research (2020).
3. Conte, M. N., Gordon, M. & Sims, C. Quarantine fatigue thins fat-tailed coronavirus impacts in US cities by making epidemics inevitable. *medRxiv* (2021).
4. Wirtz, K. W. Decline in mitigation readiness facilitated second waves of SARS-CoV-2. *medRxiv* (2021).
5. Shearston, J. A., Martinez, M. E., Nunez, Y. & Hilpert, M. Social-distancing fatigue: evidence from real-time crowd-sourced traffic data. *Sci. The Total. Environ.* 148336 (2021).
6. Giuliano, P., Rasul, I., Ciccone, A. & Ismailov, A. Compliance with social distancing during the Covid-19 crisis. *IZA Discuss. Pap.* **13114** (2020).
7. Chang, S. *et al.* Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021).
8. Templeton, A., Drury, J. & Philippides, A. From mindless masses to small groups: conceptualizing collective behavior in crowd modeling. *Rev. Gen. Psychol.* **19**, 215–229 (2015).
9. Neville, F. G. & Reicher, S. D. Crowds, social identities, and the shaping of everyday social relations. *Polit. psychology: A social psychological approach* 231–252 (2018).
10. Van Bavel, J. J. *et al.* Using social and behavioural science to support COVID-19 pandemic response. *Nat. human behaviour* **4**, 460–471 (2020).
11. Dezechache, G., Frith, C. D. & Deroy, O. Pandemics and the great evolutionary mismatch. *Curr. Biol.* **30**, R417–R419 (2020).
12. Thompson, R. N. *et al.* Key questions for modelling COVID-19 exit strategies. *Proc. Royal Soc. B* **287**, 20201405 (2020).
13. Ruktanonchai, N. W. *et al.* Assessing the impact of coordinated COVID-19 exit strategies across Europe. *Science* **369**, 1465–1470 (2020).
14. Holme, P. & Saramäki, J. Temporal networks. *Phys. reports* **519**, 97–125 (2012).
15. Read, J. M., Eames, K. T. & Edmunds, W. J. Dynamic social networks and the implications for the spread of infectious disease. *J. The Royal Soc. Interface* **5**, 1001–1007 (2008).
16. Stehlé, J., Barrat, A. & Bianconi, G. Dynamical and bursty interactions in social networks. *Phys. review E* **81**, 035101 (2010).
17. Rocha, L. E. & Blondel, V. D. Bursts of vertex activation and epidemics in evolving networks. *PLoS Comput. Biol* **9**, e1002974 (2013).
18. Hiraoka, T., Masuda, N., Li, A. & Jo, H.-H. Modeling temporal networks with bursty activity patterns of nodes and links. *Phys. Rev. Res.* **2**, 023073 (2020).

19. Girvan, M. & Newman, M. E. Community structure in social and biological networks. *Proc. national academy sciences* **99**, 7821–7826 (2002).
20. Fortunato, S. Community detection in graphs. *Phys. reports* **486**, 75–174 (2010).
21. Mucha, P. J., Richardson, T., Macon, K., Porter, M. A. & Onnela, J.-P. Community structure in time-dependent, multiscale, and multiplex networks. *Science* **328**, 876–878 (2010).
22. Reichardt, J., Alamino, R. & Saad, D. The interplay between microscopic and mesoscopic structures in complex networks. *PLoS One* **6**, e21282 (2011).
23. Jeub, L. G., Balachandran, P., Porter, M. A., Mucha, P. J. & Mahoney, M. W. Think locally, act locally: Detection of small, medium-sized, and large communities in large networks. *Phys. Rev. E* **91**, 012821 (2015).
24. Stegehuis, C., Van Der Hofstad, R. & Van Leeuwen, J. S. Epidemic spreading on complex networks with community structures. *Sci. reports* **6**, 1–7 (2016).
25. Lippold, D. *et al.* Spatiotemporal modeling of first and second wave outbreak dynamics of COVID-19 in Germany. *Biomech. modeling mechanobiology* 1–15 (2021).
26. Battiston, F. *et al.* The physics of higher-order interactions in complex systems. *Nat. Phys.* **17**, 1093–1098 (2021).
27. Truszkowska, A. *et al.* High-resolution agent-based modeling of COVID-19 spreading in a small town. *Adv. theory simulations* **4**, 2000277 (2021).
28. Mistry, D. *et al.* Inferring high-resolution human mixing patterns for disease modeling. *Nat. communications* **12**, 1–12 (2021).
29. Wing, C., Simon, D. H. & Carlin, P. Effects of large gatherings on the COVID-19 epidemic: evidence from professional and college sports. *SSRN 3657625* (2020).
30. Mangrum, D. & Niekamp, P. Jue insight: College student travel contributed to local COVID-19 spread. *J. Urban Econ.* 103311 (2020).
31. Dave, D., McNichols, D. & Sabia, J. J. The contagion externality of a superspreading event: The Sturgis Motorcycle Rally and COVID-19. *South. economic journal* **87**, 769–807 (2021).
32. Goscé, L. & Johansson, A. Analysing the link between public transport use and airborne transmission: mobility and contagion in the London underground. *Environ. Heal.* **17**, 1–11 (2018).
33. Bailey, M., Farrell, P., Kuchler, T. & Stroebel, J. Social connectedness in urban areas. *J. Urban Econ.* **118**, 103264 (2020).
34. Harris, J. E. The subways seeded the massive coronavirus epidemic in New York City. *NBER working paper* (2020).
35. Qian, X., Sun, L. & Ukkusuri, S. V. Scaling of contact networks for epidemic spreading in urban transit systems. *Sci. reports* **11**, 1–12 (2021).
36. Barrat, A. & Cattuto, C. Temporal networks of face-to-face human interactions. In *Temporal Networks*, 191–216 (Springer, 2013).
37. Sekara, V., Stopczynski, A. & Lehmann, S. Fundamental structures of dynamic social networks. *Proc. national academy sciences* **113**, 9977–9982 (2016).
38. Cremonini, M. & Maghool, S. The unknown of the pandemic: An agent-based model of final phase risks. *J. Artif. Soc. Soc. Simul.* **23** (2020).
39. Cadoni, M. & Gaeta, G. Size and timescale of epidemics in the SIR framework. *Phys. D: Nonlinear Phenom.* **411**, 132626 (2020).
40. Bandura, A. & Walters, R. H. *Social learning theory*, vol. 1 (Englewood cliffs Prentice Hall, 1977).
41. Endsley, M. R. Toward a theory of situation awareness in dynamic systems. *Hum. factors* **37**, 32–64 (1995).

Author contributions statement

M.C. and S.M. conceived the model and simulations, S.M. conducted the simulations, M.C. analyzed the results and wrote the manuscript. All authors reviewed the manuscript.

Additional information

Accession codes: Data are available at <https://zenodo.org/record/5626804>. **Competing interests:** The authors declare no competing interests.

Supplementary Information for:

The dynamical formation of ephemeral groups on networks and their effects on epidemics spreading

Marco Cremonini^{1,*} and Samira Maghool^{2,*}

¹University of Milan, Department of Political and Social Sciences, Milan, Italy

²University of Milan, Department of Computer Science, Milan, Italy

*marco.cremonini@unimi.it

*samira.maghool@unimi.it

S1: Network information and metrics

The networks used by our models for experiments have a constant population of $N=10000$ and scale-free characteristics. They are generated through the `powerlaw_cluster_graph` function of the NetworkX package (https://networkx.github.io/documentation/networkx-1.10/reference/generated/networkx.generators.random_graphs.powerlaw_cluster_graph.html), which is based on Holme and Kim algorithm¹ for graphs with powerlaw degree distribution.

Figure S1 shows an example of degree distribution, while Table S1 presents some network metrics.

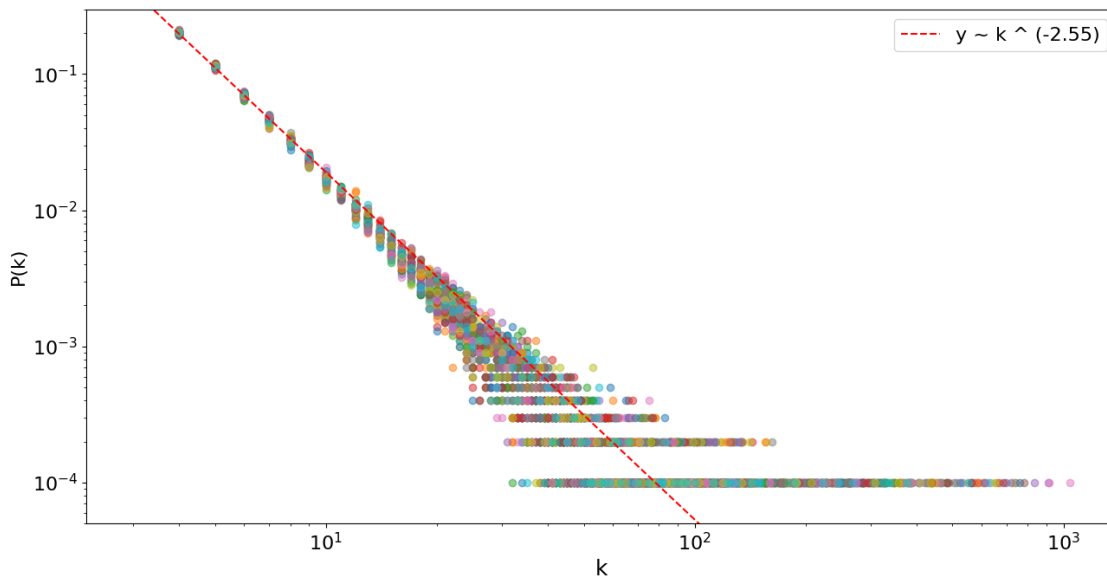


Figure S1. Degree distribution produced by trials used in simulating a single model configuration. Points of different colors correspond to different trials. The figure is on log-log scale, the red dashed line is the powerlaw fit. The distribution is typical of scale-free networks.

Table S1. Network centrality metrics.

Metric	Value
Average degree	5.99732
Average betweenness (normalized)	0.0003349
Average eigenvector	0.0035467
Average clustering	0.327954
Average Path length	4.3024

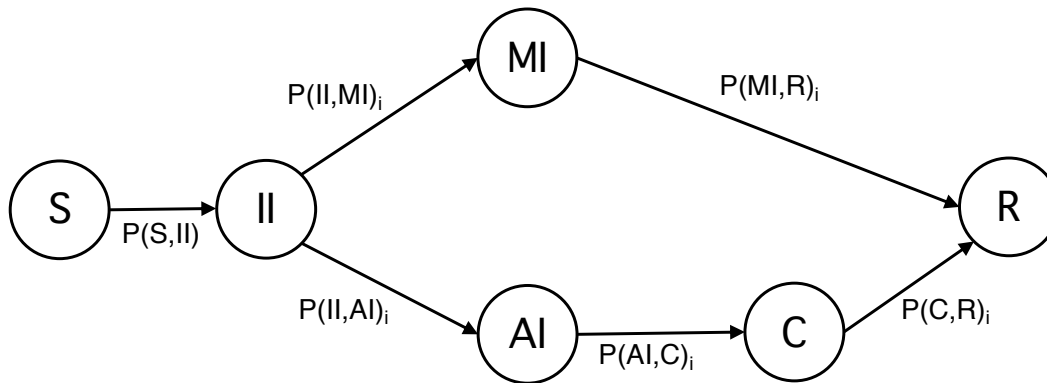
S2: Model states and execution

This section is divided in two subsections, respectively aimed at presenting:

- Model states, the state transition diagram, and pseudocode for model execution related to the epidemic spreading. The material of this subsection is not part of the original contribution of the present study, being taken from our previous work on a related subject². In that first work, we have defined the original epidemic model and network that we have used in this work for the static contact network. This model produces the first epidemic wave (see Figure 2 in the manuscript), up to the start of ephemeral grouping, and the following basic spreading mechanism on top of which ephemeral groups are created. We include this here for completeness and to ease the understanding.
- Pseudocode for ephemeral groups creation.

Model states, transition diagram, and execution pseudocode

Figure S2 shows model states and transitions. Several simplifying assumptions are made. Propagation is carried out as simple contagion, i.e., purely probabilistic state change when a susceptible agent has a direct link to an agent in an infected state. The model represents a variation of the conventional SIR, with the Infected state decomposed in subclasses: II (Incubating Infected), MI (Mild Infected), AI (Acute Infected), C (Contained). These states differ in the probability of contagion (nominal rate for AI, reduced for II and MI, zero for C) and the time spent by agents in the state (short for II and AI, longer for MI and C)^{3,4}. These subclasses could, more generally, be considered as different subpopulations of an heterogeneous class of agents with spreading ability but different degrees of effectiveness and duration. The population of agents is constant, at start up all agents have initial state S (Susceptible), with the exception of the seed agents having state II (Incubating Infected), and state R (Recovered) is final. The list of parameter settings is showed in Table S2.



Labels on links represent corresponding transition probabilities from one state to the another.

Figure S2. Model states and transition diagram.

These model assumptions represent a closed system able to produce a SIR-type dynamic (growth-peak-decline), with at most small local stochastic fluctuations, and a final stable state reached in finite time with the complete expiration of the epidemic (zero Infected agents). No oscillatory behavior (e.g., epidemic relapses, multiple epidemic waves, waves with multiple peaks) or stationary states (endemic state) are possible. Among the many aspects that are not considered, some of the most relevant are: a variable population with influx and outflux, a temporal decay of the immunity, the possibility of exogenous events, a finer definition of agent states and transitions. However, for all the improved descriptive ability that an enriched model

may provide, the increased complexity would have complicated the study aimed at introducing a variant of a SIR-type dynamic with ephemeral groups as basic mesoscopic constructs. We believe that only in a second phase of the research, an extended model should be employed by considering a richer set of characteristics. For the purpose of this study, it is also not relevant to specifically consider the different infectious states (II, MI, AI, C). What matters is that overall the basic epidemic dynamic exhibits the characteristics of a SIR-type model, therefore, with no loss of generality, in this work we consider agents in states II, MI, AI, C as aggregated in the generic class I (Infected).

The execution of the original epidemic network model, representing here the dynamic on the static contact network, is described in Algorithm 1. Each iteration represents a time step in simulation time. At every time step, each agent is selected in random order and, if in state S, its state is checked with respect to peers, or if in other states, according to time periods specific of states II, MI, AI, and C.

Ephemeral groups creation pseudocode

The execution of the ephemeral groups model is described in Algorithm 2 and Algorithm 3. Iterations starts when a given condition on the epidemic dynamic is met, corresponding to a certain time step t_G . In our case, we use a condition on the number of agents in class C during the declining phase of the first wave. The rationale for using the class C is discussed in the previous work² and it was meant to set a condition on an observable variable and by convention assume it as the start of reopening. For simplicity, we have reused the mechanism in this work too, but a corresponding condition on the aggregate class I would have equally served the same goal. In this work, the rationale is that, at a conventional level of infected agents during the declining phase, ephemeral groups start appearing. Changing the level at which ephemeral groups start changes the initial conditions of the grouping process (i.e., number agents in state S, I, R at t_G). In this work, we choose empirically the condition for starting ephemeral groups as one low enough to let us test a large variety of groups configurations, but not so low to make valid trials the exception with respect to invalid ones. In Table S2, the starting conditions of ephemeral groups are showed as average number of Susceptible, Infected and Recovered agents with 95% CI.

Table S2. Base epidemic and ephemeral groups simulation settings.

Epidemic model	
Parameter	Values
Network size	10000
Seed agents	5
Time steps (max)	300
Probability of transmission	AI state: 0.03 II & MI states: 0.015 C state: 0.0
Length of time in a state	II state: $t=[2,14]$, mean=8 MI state: $t=[2,7]$, mean=4.5 AI state: $t=[2,7]$, mean=3 C state: $t=[14,30]$, mean=22 S & R states: undefined
Proportion AI:MI	10:90
Ephemeral Groups model	
Starting conditions at t_G	Values (mean [95% CI])
Infected agents	140 [142.25,137.75]
Susceptible agents	7087 [7122.24,7051.76]
Recovered agents	2773 [2806.25,2739.85]

Algorithm 1 Epidemic network model execution

Require: Adjacency matrix $(A_{i,j})$, random seeds, parameters initialization

```
1: for t in MaxTimesteps do
2:   for i in  $A_{i,j}$  do
3:     At each time step t, for all agents in  $A_{i,j}$ , run the model according to the current agent's state
4:     Case S:
5:     if  $Node_i$  in state S then
6:       for j in  $A_{i,j} = 1$  do
7:         if  $Node_j$  in state AI then
8:           Change state to II with probability  $P(S, AI)$ 
9:         end if
10:        if  $Node_j$  in state MI then
11:          Change state to II with probability  $P(S, MI)$ 
12:        end if
13:        if  $Node_j$  in state II then
14:          Change state to II with probability  $P(S, II)$ 
15:        end if
16:      end for
17:    end if
18:    Case II:
19:    if  $Node_i$  in state II then
20:      Remain in II for  $T_{II}(i) = \text{gamma}(\alpha_1, \text{mean}_1 / \alpha_1)$  steps
21:      if  $\text{rand}[0, 1] < MI / (MI + AI)$  then
22:        Change state to MI
23:      else
24:        Change state to AI
25:      end if
26:    end if
27:    Case MI:
28:    if  $Node_i$  in state MI then
29:      Remain in MI for  $T_{MI}(i) = \text{norm}(T_{MI})$  steps
30:      When  $T_{MI}(i)$  expires:
31:      if  $\text{rand}[0, 1] < P(MI, C)$  then
32:        Change state to C
33:      else
34:        Change state to R
35:      end if
36:    end if
37:    Case AI:
38:    if  $Node_i$  in state AI then
39:      Remain in AI for  $T_{AI}(i) = \text{gamma}(\alpha_2, \text{mean}_2 / \alpha_2)$  steps
40:      When  $T_{AI}(i)$  expires, change state to C
41:    end if
42:    Case C:
43:    if  $Node_i$  in state C then
44:      if  $Node_i(t-1)$  changed state from MI then
45:        Remain in C for  $T_{C|MI}(i) = \text{norm}(T_{C|MI})$  steps
46:      else if  $Node_i(t-1)$  changed state from AI then
47:        Remain in C for  $T_{C|AI}(i) = \text{norm}(T_{C|AI})$  steps
48:      end if
49:    end if
50:    Case R:
51:    if  $Node_i$  in state R then
52:      Remain in R
53:    end if
54:  end for
55: end for
```

Algorithm 2 Ephemeral groups model execution.

Require: Ephemeral groups starting conditions matched at time step t_G of the epidemic network model execution (Algorithm 1), in the declining phase of the SIR-type dynamic

Require: Adjacency matrix $(A_{i,j})$

Require: y_1 : group size, y_2 : number of dynamic links per group, y_3 : grouping pattern, $\#G$: number of groups

```
1:  $t_0 \leftarrow t_G$ 
2: Grouping pattern is verified from parameter  $y_3$ 
3: Possible cases: Consecutive Bounded ( $y_3 : n = \bar{n}$ ), Consecutive Unbounded ( $y_3 : f = 1/1$ ), Periodic Unbounded ( $y_3 : f = 1/k$ )
4: Case 1: Consecutive Bounded
5: if  $y_3 : n = \bar{n}$  then
6:   for q in  $(1, \bar{n})$  do
7:     CreateGroups( $A_{i,j}, y_1, y_2, y_3, \#G$ )
8:   end for
9: end if
10: Case 2: Consecutive Unbounded
11: if  $y_3 : f = 1/1$  then
12:   for q in  $(1, MaxTimesteps - t_G)$  do
13:     CreateGroups( $A_{i,j}, y_1, y_2, y_3, \#G$ )
14:   end for
15: end if
16: Case 3: Periodic Unbounded
17: if  $y_3 : f = 1/k$  then
18:   for q in  $(1, MaxTimesteps - t_G)$  with step  $k$  do
19:     CreateGroups( $A_{i,j}, y_1, y_2, y_3, \#G$ )
20:   end for
21: end if
```

Algorithm 3 CreateGroup function.

```
1: function CreateGroups( $A_{i,j}$ ,  $y_1$ ,  $y_2$ ,  $y_3$ , # $G$ )
2: for  $g$  in # $G$  do
3:   Randomly select  $y_1$  agents
4:   for  $p$  in  $y_1$  do
5:      $Grp = [append(rand[1,N])]$ 
6:   end for
7:   Randomly create  $y_2$  links among selected agents
8:   for  $p$  in  $y_2$  do
9:      $x_1 = rand[Grp]; x_2 = rand[Grp]$ 
10:    Update the adjacency matrix and possible spread of contagion
11:     $A_{x_1,x_2} = 1$ 
12:    if  $x_1$  in state S then
13:      if  $x_2$  in state AI then
14:        Change state to II with probability  $P(S,AI)$ 
15:      end if
16:      if  $x_2$  in state MI then
17:        Change state to II with probability  $P(S,MI)$ 
18:      end if
19:      if  $x_2$  in state II then
20:        Change state to II with probability  $P(S,II)$ 
21:      end if
22:    else if  $x_2$  in state S then
23:      if  $x_1$  in state AI then
24:        Change state to II with probability  $P(S,AI)$ 
25:      end if
26:      if  $x_1$  in state MI then
27:        Change state to II with probability  $P(S,MI)$ 
28:      end if
29:      if  $x_1$  in state II then
30:        Change state to II with probability  $P(S,II)$ 
31:      end if
32:    end if
33:  end for
34:  Remove dynamic links  $A_{\bar{i},\bar{j}} = 0$ 
35: end for
```

S3: Null Model analysis

In this section, we present a *null model analysis*, testing the expectation that any patterns in simulation outcomes arise only from random sampling processes (i.e., *null hypothesis*)^{5,6}. The *alternative hypothesis* is that patterns in data are not random variations produced by the null hypothesis and therefore assuming the creation of ephemeral groups as relevant for the outcome cannot be rejected. For the analysis, we assume as data the time series with the average number of agents in state Infected, calculated with respect to the simulation trials. Averaging is preceded by the realignment of trials with respect to the flex's time step, which represents the beginning of the second epidemic dynamics. This step has been commented in the manuscript. In the following, we use the expression *Group model* to refer to our original network model based on ephemeral groups.

Random Dynamic Links (RDL) and Constant Random Dynamic Links (CRDL) null models

Two null models have been defined:

- **Random Dynamic Links (RDL):** Dynamic links are created randomly in each time step of a series of dynamic linking time steps, with agents not organized in groups. A RDL null model is compared to a Group model with the constraint that for each time step of a series of dynamic linking time steps, the total number of random links created in the null model is equal to the sum of links created randomly in all groups of the Group model.
- **Constant Random Dynamic Links (CRDL):** This is a variation of the RDL null model still having random dynamic links created in the first time step of a series of dynamic linking time steps, but then kept unchanged in the following time steps.

Given the general description of the null models, the competing hypothesis could be stated as follows:

- H_0 : *Null Hypothesis*: The time series of the Infected agents resulting from a Group model could also be explained as resulting from a corresponding RDL or CRDL null model.
- H_1 : *Alternative Hypothesis*: The time series of the Infected agents resulting from a Group model cannot be explained as resulting from one of the null models.

Non-nested model

Before presenting test results, few considerations are needed regarding the rival statistical models we defined for the Group model and for the RDL/CRDL null models in order to perform a significance test. In particular, we defined them as *non-nested models*, meaning that the RDL/CRDL null models are not included in the Group model^{7,8}.

Formally, the general representation of a non-nested model is a joint probability density function:

$$H_f : f(\mathbf{y}_t | \theta), \theta \in \Theta$$

$$H_g : g(\mathbf{y}_t | \gamma), \gamma \in \Gamma$$

with $f(\mathbf{y}_t | \theta)$ and $g(\mathbf{y}_t | \gamma)$ two density functions, θ and γ unknown but admissible parameters in parameter spaces Θ and Γ .

In the special case of linear regression models, the one we adopted, the form of the non-nested model is:

$$H_f : \mathbf{y} = \mathbf{X}\alpha + \mathbf{u}_f$$

$$H_g : \mathbf{y} = \mathbf{Z}\beta + \mathbf{u}_g$$

with \mathbf{X} and \mathbf{Z} two vectors of observations on the explanatory variables of models H_g and H_f , u_f and u_g as noise factors with normal distribution, and α and β the unknown regression coefficients⁸.

Through a non-nested model, the goal is to evaluate which of the rival models, H_g and H_f , can be considered the best predictor, if any, of the unknown joint probability density function, $f_0(\mathbf{y})$. Therefore, in its general form, the alternatives in a non-nested model simply represent different theories, none of which necessarily has the characteristics of a null model. This is also the reason for the different notation, H_g and H_f instead of H_0 and H_1 .

However, since in our case we do test our Group model with respect to the RDL/CRDL null models, for conformity we maintain the traditional H_0 and H_1 notation.

Explanatory variables and linear regression models

Our aim is to test the alternative hypothesis with respect to data represented by the time series with the average number of agents in state Infected. The logic we followed in the definition of the models consider that the algorithm for dynamic link creation is what differs between the hypothesis. Therefore explanatory variables should be defined to reflect the effect of those

different algorithms. By design, the number of dynamic links per time step is not an explanatory variable, because for the alternative models that number is the same. The number and frequency of time steps with dynamic links is defined by the specific grouping pattern (e.g., limited or unlimited consecutive time steps, periodic with a certain frequency). However, there is a key difference between the Group model and the RDL/CRDL null models in the number of dynamic links that *could propagate a contagion*, which are only those connecting a Susceptible agent on one end and an Infected agent at the other end. For simplicity, we refer to these links as *Active links*. Here is the main observation of our hypothesis testing:

The number of Active links, among the total number of dynamic links created in a time step, depends on the different dynamic link creation algorithms of Group model and RDL/CRDL null models, therefore it can be used as explanatory variable for the non-nested models.

Consequently, the linear regression models could be specified as:

$$\begin{array}{lll} \text{Null Model} & H_0 : \text{Infected} \sim \text{ActiveLinks_RDLnull} & H_0 : \text{Infected} \sim \text{ActiveLinks_CRDLnull} \\ \text{Group Model} & H_1 : \text{Infected} \sim \text{ActiveLinks_AllGroups} & H_1 : \text{Infected} \sim \text{ActiveLinks_AllGroups} \end{array}$$

with *Infected* the time series of Infected agents produced by the Group model, and explanatory variables *ActiveLinks_RDLnull*, *ActiveLinks_CRDLnull*, and *ActiveLinks_AllGroups* corresponding to the time series of Active links produced in the three models.

Number of Active links

The probability for a random link to be an Active link for each timestep can be estimated from model simulations. Statistics needed for the evaluation are: *rate of Infected agents* $P_t(\text{Infected})$ and *rate of Susceptible agents* $P_t(\text{Susceptible})$ per time step. From these, the probability $P_t(\text{ActiveLink})$ that, at a certain time step, a dynamic link is an Active link is:

$$P_t(\text{ActiveLink}) = 2 * P_t(\text{Infected}) * P_t(\text{Susceptible})$$

Next, we have estimated the number of *unique links* created per time step. This number is different from the total number of dynamic links created per time step, because dynamic links can be repeated.

The number of unique links, resulting from the random selection with repetition of k links over n possible links, can be evaluated by:

$$\text{UniqueLinks} = n \left(1 - \left(1 - \frac{1}{n} \right)^k \right)$$

which results, for the different models:

$$\begin{array}{ll} \text{Group Model} & \text{UniqueLinks_singlegroup} = a \left(1 - \left(1 - \frac{1}{a} \right)^{y_2} \right) \quad \text{and} \quad a = \frac{y_1(y_1 - 1)}{2} \\ \text{RDL/CRDL Models} & \text{UniqueLinks_global} = b \left(1 - \left(1 - \frac{1}{b} \right)^{\#G * y_2} \right) \quad \text{and} \quad b = \frac{N(N - 1)}{2} \end{array}$$

with, for the Group model, a group of size y_1 and number of dynamic links per time step y_2 , while for null models, $\#G * y_2$ represents the total random links created over the entire population N . The following important property holds:

$$\text{UniqueLinks_global} \neq \#G * \text{UniqueLinks_singlegroup}$$

Finally, the *number of Active links* at each time step can be calculated as:

$$\begin{array}{ll} \text{Group Model} & \text{ActiveLinks_AllGroups}_t = \#G * \text{UniqueLinks_singlegroup} * P_t(\text{ActiveLinks}|\text{Group model}) \\ \text{RDL Model} & \text{ActiveLinks_RDLnull}_t = \text{UniqueLinks_global} * P_t(\text{ActiveLinks}|\text{RDL model}) \\ \text{CRDL Model} & \text{ActiveLinks_RDLnull}_t = \text{UniqueLinks_global} * P_t(\text{ActiveLinks}|\text{CRDL model}) \end{array}$$

with the notation $P_t(\text{ActiveLinks}|\dots\text{model})$ indicating the probability that a random link is an Active link at time step t given Group, RDL, or CRDL model.

Hypothesis testing

With the definition of non-nested model and the three time series of Active links as explanatory variables, resulting from trials of Group model and RDL/CRDL null models simulations, we have been able to test the hypothesis.

As statistical testing methodology, we selected the *Encompassing test* being well-suited for non-nested models^{7,9,10}.

This test is available as a convenient R function¹¹ that takes two non-nested linear regression models as input, generically called *Model 1* and *Model 2*, and a data frame containing the time series. Operationally, it defines a new model called *Model E* as the one composed by the explanatory variables of Model 1 and Model 2, so that both original models are nested in the new encompassing model. Then it proceeds to test Model 1 Vs. Model E and Model 2 Vs. Model E. The method used is Wald test for nested models, similar to ANOVA test¹².

The output produces the *F* statistics of the two tests with the corresponding significance level and degrees of freedom. Checking a *F* distribution table at a certain significance level (alpha level), first it could be verified if there is a statistically significant difference between the mean scores of Model 1 or Model 2 with respect to the encompassing Model E. If a significant difference is confirmed, then the *F* statistics obtained by Model 1 and Model 2 could be compared and that with the higher value has the better explanatory power. For example, if Model 1 corresponds to the Group model and Model 2 to the RDL null model, we will have:

Group Model	Model 1 : <i>Infected</i> ~ <i>ActiveLinks_AllGroups</i>
Null Model	Model 2 : <i>Infected</i> ~ <i>ActiveLinks_RDLnull</i>
Encompassing Model	Model E : <i>Infected</i> ~ <i>ActiveLinks_AllGroups</i> + <i>ActiveLinks_RDLnull</i>

The outcome will have a form similar to:

	Res.Df	Df	F	Pr(>F)
M1 vs. ME	160	-1	309.60	< 2.2e-16 ***
M2 vs. ME	160	-1	189.38	< 2.2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

with *Res.Df* and *Df* respectively the residual and the difference of regression degrees of freedom, *F* the *F* value, and *Pr(>F)* the significance level. In the example, both models show a significant difference with respect to the encompassing model and Model 1 has a better explanatory power than Model 2. In this case, we can reject the RDL null model.

Results and discussion

Results of hypothesis testing are presented in Table S3 for a sample of the 160 configurations simulated in this work. In particular, for each scenarios defined by $\#G = (10, 100, 1000)$ and each series of simulations with same group size and number of dynamic links per group, (y_1, y_2) , we selected for testing the configuration that first reached the peak of Infected. Configurations that plateaued reached the peak in a limited number of time steps (e.g. $y_3 : n = 12$ or $y_3 : n = 16$), others reached the peak with the continuous unlimited pattern ($y_3 : f = 1/1$) in a number of time steps greater of 16. The scenario with $\#G = 1$ is not considered because only a single group of agents is created with random links, therefore it is equivalent to random models. In Table S3, the color code should give an immediate feeling of the test results. A red square means that the null model cannot be rejected; a blue square that the Group model has a better statistical significance. When the *F* statistics were close we signal it adding “weak”.

Results from Table S3 look interesting, we believe, because they show a pattern representing a change of significance of the Group model with respect to random models following the increase of the group size. For the smallest groups there is no statistical difference between the null models and the Groups model; basically small ephemeral groups are effective just because they create new temporal connections, being the group structure unimportant. Then, by increasing the group size, the difference progressively emerges, first weakly, then more evident. For mid-sized to large groups, the fact they have a group structure (even non persistent or temporary, as in our cases) becomes relevant.

This evidence, may have interesting logical explanations with respect to the group behavior, not just between very small and larger groups, but also considering a dynamics of groups, for example starting as small ephemeral groups and progressively forming larger social structures. A conclusion could be that the Null Model analysis points to a relevance of mesoscopic constructs that varies with changes in social organization, groups nature, formation, and aggregation, and that can be included and accounted for in agent-group network models.

The two random models gave same testing results, with respect to the Group model. We suppose that this is due to the fact that Active links are in any case in small numbers and sparse with respect to the whole network, and new Infected agents directly created by Active links are a large minority with respect to the total (similarly to seed nodes with respect to the Infected produced by the epidemic). Therefore, maintaining the same dynamic links, as in the CRDL null model, or recreating anew as

in the RDL null model does not make much difference, in our context. Anyhow, this does not mean that it could not be relevant under different conditions and models or that a similar variant applied to the Group model (i.e., creating persistent groups that maintain constant their members for more timesteps) is not worth studying. We think they could produce interesting results and be the subject of future works.

Table S3. Hypothesis testing: Group model Vs. RDL null model and Group model Vs. CRDL null model. For each test, the alternative with higher F value is showed in bold. Colored boxes help to interpret testing results: **Red** when the Null model has better explanatory power, **Blue** when the Group model has. The notation "weak" is qualitative and indicates when F values between Group and Null models are close. **Parameters:** #G is the number of groups created per time step, y_1 the group size, y_2 the number of random links per group and per time step, and y_3 the grouping pattern (i.e., a finite number of time steps n or the consecutive unlimited case $f = 1/1$).

Parameters	Model 1: Groups - Model 2: RDL null		Model 1: Groups - Model 2: CRDL null	
#G=1000	Model 1: Infected ~ActiveLinks_AllGroups Model 2: Infected ~ActiveLinks_RDLnull Model E: Infected ~ActiveLinks_AllGroups + ActiveLinks_RDLnull — Test - Residual Df - Regression Df - F - Pr(>F) Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1		Model 1: Infected ~ActiveLinks_AllGroups Model 2: Infected ~ActiveLinks_CRDLnull Model E: Infected ~ActiveLinks_AllGroups + ActiveLinks_CRDLnull — Test - Residual Df - Regression Df - F - Pr(>F) Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1	
	y1= 20; y2=100; y3: f=1/1	M1 vs. ME 170 -1 460.66 <2.2e-16 *** M2 vs. ME 170 -1 776.54 <2.2e-16 ***		M1 vs. ME 170 -1 285.07 <2.2e-16 *** M2 vs. ME 170 -1 449.61 <2.2e-16 ***
y1= 20; y2=200; y3: n=16	M1 vs. ME 167 -1 9.0464 0.0030394 ** M2 vs. ME 167 -1 13.0051 0.0004099 ***	weak	M1 vs. ME 167 -1 16.193 8.657e-05 *** M2 vs. ME 167 -1 21.223 8.075e-06 ***	weak
y1= 50; y2=250; y3: n=12	M1 vs. ME 169 -1 5.3465 0.02197 * M2 vs. ME 169 -1 4.8956 0.02827 *	weak	M1 vs. ME 169 -1 7.9032 0.005518 ** M2 vs. ME 169 -1 7.1492 0.008236 **	weak
y1= 50; y2=500; y3: f=1/1	M1 vs. ME 167 -1 13.6463 0.0002987 *** M2 vs. ME 167 -1 9.8225 0.0020368 **	weak	M1 vs. ME 167 -1 5.5260 0.01990 * M2 vs. ME 167 -1 3.9671 0.04803 *	weak
#G=100				
y1= 50; y2=500; y3: f=1/1	M1 vs. ME 160 -1 3500.6 <2.2e-16 *** M2 vs. ME 160 -1 1161.3 <2.2e-16 ***		M1 vs. ME 160 -1 309.60 <2.2e-16 *** M2 vs. ME 160 -1 189.38 <2.2e-16 ***	
y1= 50; y2=1000; y3: f=1/1	M1 vs. ME 168 -1 8.5678 0.003896 ** M2 vs. ME 168 -1 0.7394 0.391064		M1 vs. ME 168 -1 178.111 <2.2e-16 *** M2 vs. ME 168 -1 78.296 1.19e-15 ***	
y1= 50; y2=2000; y3: n=16	M1 vs. ME 167 -1 37.658 5.924e-09 *** M2 vs. ME 167 -1 13.593 0.0003067 ***		M1 vs. ME 167 -1 37.417 6.554e-09 *** M2 vs. ME 167 -1 14.582 0.0001888 ***	
y1= 100; y2=1000; y3: f=1/1	M1 vs. ME 169 -1 312.76 <2.2e-16 *** M2 vs. ME 169 -1 154.90 2.2e-16 ***		M1 vs. ME 169 -1 697.57 <2.2e-16 *** M2 vs. ME 169 -1 265.07 2.2e-16 ***	
y1= 100; y2=2000; y3: n=16	M1 vs. ME 170 -1 45.230 2.554e-10 *** M2 vs. ME 170 -1 23.151 3.286e-06 ***		M1 vs. ME 170 -1 45.952 1.910e-10 *** M2 vs. ME 170 -1 25.026 1.401e-06 ***	
y1= 200; y2=2000; y3: n=16	M1 vs. ME 169 -1 38.319 4.405e-09 *** M2 vs. ME 169 -1 30.544 1.216e-07 ***		M1 vs. ME 169 -1 38.646 3.841e-09 *** M2 vs. ME 169 -1 31.984 6.506e-08 ***	
#G=10				
y1= 100; y2=6000; y3: f=1/1	M1 vs. ME 167 -1 4517.34 <2.2e-16 *** M2 vs. ME 167 -1 373.77 <2.2e-16 ***		M1 vs. ME 167 -1 2558.29 <2.2e-16 *** M2 vs. ME 167 -1 730.65 <2.2e-16 ***	
y1= 200; y2=8000; y3: f=1/1	M1 vs. ME 173 -1 2264.140 <2.2e-16 *** M2 vs. ME 173 -1 36.645 8.563e-09 ***		M1 vs. ME 173 -1 7443.28 <2.2e-16 *** M2 vs. ME 173 -1 382.17 <2.2e-16 ***	
y1= 500; y2=20000; y3: n=16	M1 vs. ME 169 -1 60.707 6.407e-13 *** M2 vs. ME 169 -1 14.042 0.000245 ***		M1 vs. ME 169 -1 80.234 5.907e-16 *** M2 vs. ME 169 -1 11.239 0.0009884 ***	

S4: Non Overlapping Group model

Another model analysis has been conducted in order to evaluate the statistical difference between the Group model and a variant with the additional constraint of creating only *disjoint groups* in the same time step. We refer to this case as the *Non Overlapping Group model*. The aim of the analysis and the way we proceed are similar to what presented in the previous Null Model analysis section.

The main difference is that, while we still have non-nested linear regression models with explanatory variables represented by the Active links produced in the two cases, differently from the Null Model analysis, now none of the alternatives has the characteristics of a null model. Rather, they are models based on alternative algorithms (i.e., random choice with repetition Vs. random choice without repetition), representing different theories.

Operationally, this requires testing models reciprocally. Therefore, two set of hypothesis are tested.

Case A:

- $H_{Overlap}^A$: The time series of Infected agents produced by the Group model (*Infected_Overlapping*) cannot be explained as resulting from the Non Overlapping Group model.
- $H_{NonOverlap}^A$: The *Infected_Overlapping* time series can be explained as resulting from the Non Overlapping Group model.

Case B:

- $H_{NonOverlap}^B$: The time series of Infected agents produced by the Non Overlapping Group model (*Infected_NonOverlapping*) cannot be explained as resulting from the Group model.
- $H_{Overlap}^B$: The *Infected_NonOverlapping* time series can be explained as resulting from the Group model.

Correspondingly, the linear regression models become:

Case A:

$$H_{Overlap}^A : \text{Infected_Overlapping} \sim \text{ActiveLinks_AllGroups}$$

$$H_{NonOverlap}^A : \text{Infected_Overlapping} \sim \text{ActiveLinks_NonOverlapping}$$

Case B:

$$H_{NonOverlap}^B : \text{Infected_NonOverlapping} \sim \text{ActiveLinks_NonOverlapping}$$

$$H_{Overlap}^B : \text{Infected_NonOverlapping} \sim \text{ActiveLinks_AllGroups}$$

Results and discussion









Results of hypothesis testing are presented in Table S4 for a sample of four configurations with different characteristics. In general, they tend to reject the hypothesis that the dynamics produced by the two models are statistically different, meaning that for groups studied in this work, the possible overlapping does not have dominant effects. Some observations are, however, possible to discuss. For instance, the configuration with parameters ($\#G = 1000$, $y_1 = 10$, $y_2 = 50$) shows some interesting properties:

- All agents of the population ($N = 10000$) belong to an ephemeral group ($\#G * y_1 = 10000$) of size 10. Therefore this is the scenario with highest rate of agents involved in the dynamic link mechanism, but highest group fragmentation (further reducing the group size makes the epidemic unlikely to start).
- It is the case for which the Non Overlapping model has the highest explanatory power, meaning that the effect of overlapping groups is not the relevant factor.

If this result is combined with those of Table S3 showing that for small groups both RDL and CRDL null models cannot be rejected, it could be advanced the explanation that the actual effect of small ephemeral groups on the dynamics is to foster the creation of additional dynamic links, having the group structure no particular relevance with respect to both the overall dynamics (this is the observation we made in the Null Model analysis section) and the possible high-order coupling between different groups.

For the other configurations tested, no model shows a clearly better explanatory power, with F statistics very close between models, both for Case 1 and Case 2. This seems to indicate that groups overlaps is not particularly relevant in our current Group model. As for precedent explanations, this outcome is likely dependent to the limited time span of ephemeral groups, whose agents and dynamic links are recreated randomly at each grouping time steps, every time destroying group overlaps. Trying

Table S4. Model analysis: Group model Vs. Non Overlapping Group model. For each test, the alternative with higher F value is showed in bold. Colored boxes help to interpret testing results: **Blue** when Model 1 has better explanatory power, **Blue** when Model 2 has. The notation "weak" is qualitative and indicates when F values between tested models are close. **Parameters:** #G is the number of groups created per time step, y_1 the group size, y_2 the number of random links per group and per time step, and y_3 the grouping pattern (i.e., in this sample always corresponding to the consecutive unlimited case $f = 1/1$).

Parameters	Case 1. Model 1: Groups - Model 2: Non Overlapping	
	Model 1: Infected_Overlapping ~ Active_Links_AllGroups Model 2: Infected_Overlapping ~Active_Links_NonOverlapping Model E: Infected_Overlapping ~Active_Links_AllGroups + Active_Links_NonOverlapping — Residual Df - Regression Df - F value - Pr(>F) Signif. codes: 0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1	
#G=1000 y1= 10; y2=50; y3: f=1/1	M1 vs. ME 170 -1 460.66 <2.2e-16 *** M2 vs. ME 170 -1 776.54 <2.2e-16 ***	
#G=100 y1= 50; y2=1000; y3: f=1/1	M1 vs. ME 167 -1 13.6463 0.0002987 *** M2 vs. ME 167 -1 9.8225 0.0020368 **	
#G=10 y1= 200; y2=8000; y3: f=1/1	M1 vs. ME 167 -1 9.0464 0.0030394 ** M2 vs. ME 167 -1 13.0051 0.0004099 ***	
#G=10 y1= 100; y2=6000; y3: f=1/1	M1 vs. ME 169 -1 5.3465 0.02197 * M2 vs. ME 169 -1 4.8956 0.02827 *	
	Case 2. Model 1: Non Overlapping - Model 2: Groups	
	Model 1: Infected_NonOverlapping ~Active_Links_NonOverlapping Model 2: Infected_NonOverlapping ~Active_Links_AllGroups Model E: Infected_NonOverlapping ~ Active_Links_NonOverlapping + Active_Links_AllGroups — Residual Df - Regression Df - F value - Pr(>F) Signif. codes: 0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1	
#G=1000 y1= 10; y2=50; y3: f=1/1	M1 vs. ME 170 -1 449.61 <2.2e-16 *** M2 vs. ME 170 -1 285.07 <2.2e-16 ***	
#G=100 y1= 50; y2=1000; y3: f=1/1	M1 vs. ME 167 -1 3.9671 0.04803 * M2 vs. ME 167 -1 5.5260 0.01990 *	
#G=10 y1= 200; y2=8000; y3: f=1/1	M1 vs. ME 167 -1 21.223 8.075e-06 *** M2 vs. ME 167 -1 16.193 8.657e-05 ***	
#G=10 y1= 100; y2=6000; y3: f=1/1	M1 vs. ME 169 -1 7.1492 0.008236 ** M2 vs. ME 169 -1 7.9032 0.005518 **	

again to relate to the results of Table S3, other effects may have stronger influence over the dynamics in our case: firstly, the group structure, which testing results of Table S3 suggest to be increasingly relevant with increasing group size, and secondly the effect on agents outside groups but connected through the static contact network, this suggested by results discussed in the manuscript. A possible hypothesis is that these two effects overcome the one produced by groups overlaps.

At any rate, these possible explanations should be considered as educated guesses supported by a few preliminary theoretical results, at best. As a matter of fact, groups overlaps in dynamic network models of epidemics is a feature extremely relevant and still not well understood. Disjoint groups and randomly overlapping groups are two instances of a continuum described by a variable degree of overlapping among groups and temporal patterns of overlapping. As a subject for future research, group overlapping promises interesting challenges and relevant insights, we believe.

References

1. Holme, P. & Kim, B. J. Growing scale-free networks with tunable clustering. *Phys. review E* **65**, 026107 (2002).
2. Cremonini, M. & Maghool, S. The unknown of the pandemic: An agent-based model of final phase risks. *J. Artif. Soc. Soc. Simul.* **23** (2020).
3. Keeling, M. J. & Eames, K. T. Networks and epidemic models. *J. Royal Soc. Interface* **2**, 295–307 (2005).
4. Brauer, F. Compartmental models in epidemiology. In *Mathematical epidemiology*, 19–79 (Springer, 2008).
5. Gotelli, N. J. & Ulrich, W. Statistical challenges in null model analysis. *Oikos* **121**, 171–180 (2012).
6. Veech, J. A. Significance testing in ecological null models. *Theor. Ecol.* **5**, 611–616 (2012).
7. Pesaran, M. H. Non-nested hypotheses. In *Econometrics*, 167–173 (Springer, 1990).
8. Pesaran, M. H. & Weeks, M. Non-nested hypothesis testing: an overview. *A companion to theoretical econometrics* 279–309 (2001).
9. Davidson, R., MacKinnon, J. G. *et al.* Estimation and inference in econometrics. *OUP Catalogue* (1993).
10. Mizon, G. E. & Richard, J.-F. The encompassing principle and its application to testing non-nested hypotheses. *Econom. J. Econom. Soc.* 657–678 (1986).
11. R Documentation. *encomptest: Encompassing test for comparing non-nested models* (2021).
12. Fahrmeir, L., Kneib, T., Lang, S. & Marx, B. Regression models. In *Regression*, 21–72 (Springer, 2013).

,