Master equation analysis of mesoscopic localization in contagion dynamics on higher-order networks

Guillaume St-Onge,1,2 Vincent Thibeault,1,2 Antoine Allard,1,2 Louis J. Dubé,1,2 and Laurent Hébert-Dufresne1,3,4

1Département de physique, de génie physique et d’optique, Université Laval, Québec (Québec), Canada G1V 0A6
2Centre interdisciplinaire en modélisation mathématique, Université Laval, Québec (Québec), Canada G1V 0A6
3Vermont Complex Systems Center, University of Vermont, Burlington, VT 05405
4Department of Computer Science, University of Vermont, Burlington, VT 05405

Simple models of infectious diseases tend to assume random mixing of individuals, but real interactions are not random pairwise encounters: they occur within clearly defined higher-order structures such as workplace, households, schools, and concerts. We model contagion on networks with higher-order structures using clique-based approximate master equations, in which we track all states and interactions within a clique and assume a mean-field coupling between them. Using the Susceptible-Infected-Susceptible dynamics, our approach allows us to show the existence of a mesoscopic localization regime, where a disease can concentrate and self-sustain only around large substructures in the network. In this regime, the phase transition is smeared, characterized by an inhomogeneous activation of the various higher-order structures. At the mesoscopic level, we observe that the distribution of infected nodes within cliques of a same size can be very dispersed, even bimodal. When considering networks heterogeneous both at the level of nodes and cliques, we characterize analytically the region associated with mesoscopic localization in the structural parameter space. We put in perspective this phenomenon with eigenvector localization and discuss how a focus on higher-level structures is needed to discern the more subtle localization at the mesoscopic level. Finally, we discuss how mesoscopic localization affects the response to structural interventions and how this framework could provide important insights for a broad range of dynamics.

I. INTRODUCTION

Classic epidemiological models have been successful at providing meaningful insights on the spreading of infectious diseases [1,2]. Their simplicity is their strength: from as little information as the basic reproduction number \( R_0 \), one can tell whether or not a disease should invade or not a population. However, we cannot hope to represent the complexity of human behavior and of our modern social structure with mathematical models relying solely on an average individual. This is even more true when considering more complex types of spreading processes, such as social contagions [3–5] or the coevolution of diseases [6].

The study of spreading processes on networks allows to look beyond the mass action principle, to account for more realistic contact patterns while keeping our models simple enough to provide meaningful insights [7,8]. One success of network science has been to unveil the impact of the heterogeneity of contacts, and how this affects critical properties of these systems. Indeed, heterogeneous mean-field theories [9,10], heterogeneous pair approximations [11,12], and approximate master equations [13–15] represent only a few of the many techniques that have been developed to describe the behavior of dynamical processes on networks with heterogeneous number of contacts.

Social networks, however, are more than just random contacts between heterogeneous individuals: interactions occur in a coordinated manner because of a higher-level structure. At the mesoscopic level, we see groups of individuals that are more or less densely connected to one another [16,17]. We can thus shift from asking if a contagion can invade a population, to where it should thrive within that population [18].

This question is best embodied by the phenomenon of epidemic localization: near the epidemic threshold, the disease exists only in some parts of the whole network.

The localization of epidemics has been studied mostly through the lens of extensive numerical simulations or quenched mean-field theory [19–22]. A general observation is that for most complex networks, an epidemic should either be localized around the innermost network core or the hubs [20]. The localization subgraph depends on the structure, but also on the details of the dynamics [23,24]. Moreover, localization dramatically affects the fundamental critical properties of an epidemic: it is notably possible to observe a Griffiths phase, where the system slowly relaxes to an inactive state [25,29]. Another notable effect is the smearing of the phase transition, where the order parameter develops inhomogeneously beyond the critical point [27,31].

Despite the important body of work on epidemic localization, theoretical results are still limited to a handful of models and most works have used a node-centric perspective on the matter. To broaden our understanding of localization of dynamical processes on networks, we argue that approximate master equations represent powerful and flexible approaches to that end. In this paper, we present an approximate master equation analysis of what we call mesoscopic localization for the SIS dynamics on heterogeneous networks with higher-order structures (see Fig. 1). This phenomenon is characterized by the localization of the contagion in large but finite-size mesoscopic structures near the epidemic threshold [32], with a phase transition that is smeared at the global level. In this paper, we present a complete analytical description of this localization regime, while we describe its impact on interventions in Ref. [33] to show how accounting for this localization
Delocalized regime

Mesoscopic localization regime

FIG. 1. Simple illustration of the mesoscopic localization phenomenon. In both regimes, the contagion is concentrated around the innermost core of the network, but the composition of the core is different. In what we called the delocalized regime, substructures of all sizes (e.g. triangles, 4-cliques, etc.) contribute to the contagion, while in the mesoscopic localization regime, there is a strong bias toward the largest and densest substructures.

This paper is structured as follows. We first introduce a clique-based framework in Sec. II. We obtain an implicit expression and explicit bounds for the epidemic threshold in Sec. II A. With a development of the stationary state near the critical point, we show that mesoscopic localization emerges from a sufficiently weak coupling between the cliques in Sec. II B. Second, we fully characterize mesoscopic localization in Sec. III. We derive asymptotic results for the scaling of the epidemic threshold in Sec. III A leading to explicit expressions for the localization regimes. We then consider the effects of finite-size substructures in Sec. III B. We complete our analysis using the inverse participation ratio, further connecting our work with the literature on eigenvector localization. Our comparison reveals the importance of a change of perspective—a focus on higher-order structures rather than nodes—in order to detect localization phases at the mesoscopic level. Finally, in Sec. IV we discuss possible extensions of our work and some direct implications for the control of epidemics [33].

II. CLIQUE-BASED SIS MODEL

The simplest type of finite-size structure to represent a group of nodes is a clique, i.e. a fully connected subgraph. Cliques can be of different sizes and each node has a membership \( m \), corresponding to its participation to different cliques. We consider infinite-size heterogeneous random networks with cliques, characterized by a clique size distribution \( p_n \) and a membership distribution \( g_m \) [17]. Nodes are assigned to cliques uniformly at random—there are no correlations between \( m \) and \( n \). Throughout the paper, we denote expected values taken over \( p_n \) and \( g_m \) as \( \langle \cdots \rangle \), where the interior of the bracket makes it clear on which distribution the average is performed.

Let us introduce a few structural properties associated with this ensemble. The average membership of a node is \( \langle m \rangle \) and the average clique size is \( \langle n \rangle \). If we select a random node within any clique, its membership distribution is proportional to \( mg_m \). The average excess membership of this node, i.e. participation to cliques excluding the one we picked it from, is equal to \( \langle m(m - 1) \rangle / \langle m \rangle \). Similarly, the distribution of the size of a clique to which a random node belongs is proportional to \( np_n \) and the average excess clique size, i.e. number of neighbors this node has in that clique, is \( \langle n(n - 1) \rangle / \langle n \rangle \). Since \( m \) and \( n \) are uncorrelated, the average degree of a node is therefore

\[
\frac{\langle m \rangle \langle n(n - 1) \rangle}{\langle n \rangle}.
\]

On these networks, we consider the Susceptible-Infected-Susceptible (SIS) dynamics in which each node is either infected or susceptible. Infected nodes transmit the disease to their neighbors at rate \( \beta \) and recover to the susceptible state at a rate set to 1 without loss of generality. We describe the dynamics using the heterogeneous clique approximation of Ref. [34]. We track \( S_m(t) \), the fraction of nodes of membership \( m \) that are susceptible at time \( t \), and \( C_{n,i}(t) \) the fraction of cliques that are of size \( n \) with \( i \in \{0, \ldots, n\} \) infected nodes at time \( t \). It is also useful to factorize \( g_m \) and \( p_n \), yielding \( s_m(t) \equiv S_m(t)/g_m \), the probability of a node of membership \( m \) to be susceptible, and \( c_{n,i}(t) = C_{n,i}(t)/p_n \), the probability to observe \( i \) infected nodes within a clique of size \( n \).

We define the following system of approximate master equations

\[
\frac{ds_m}{dr} = 1 - s_m - mrs_m, \tag{1a}
\]

\[
\frac{dc_{n,i}}{dr} = (i + 1)c_{n,i+1} - ic_{n,i} + (n - i + 1)[\beta(i - 1) + \rho]c_{n,i-1} - (n - i)[\beta + \rho]c_{n,i}, \tag{1b}
\]

which contains a total of \( O(m_{\text{max}} + n_{\text{max}}^2) \) equations, where \( m_{\text{max}} \) and \( n_{\text{max}} \) are the maximal membership and maximal clique size respectively. The mean fields \( r(t) \) and \( \rho(t) \) are de-
fined as

\[ r(t) = \frac{\sum_{n,i} \beta(n-i)c_{n,i}(t)p_n}{\sum_{n,i} (n-i)c_{n,i}(t)p_n}, \quad (2a) \]

\[ \rho(t) = r(t) \left[ \frac{\sum_m m(m-1)s_m(t)g_m}{\sum_m m s_m(t)g_m} \right]. \quad (2b) \]

If we take a random susceptible node within a clique, \( r(t) \) is the mean infection rate associated with a random external clique to which it belongs and \( \rho(t) \) is the mean excess infection rate caused by all external cliques (excluding the one we picked the node from). The global prevalence (average fraction of infected nodes) is then

\[ I(t) = \sum_m g_m [1 - s_m(t)], \]

and the prevalence within cliques of size \( n \) is

\[ I_n(t) = \frac{1}{n} \sum_i i c_{n,i}(t). \]

Note that unless specified otherwise, sums over \( m(n) \) are over every value such that \( g_m > 0 (p_n > 0) \), and sums over \( i \) cover the range \([0, \ldots, n]\).

In Eq. (1), the evolution of each \( s_m \) is treated exactly, while the contribution of infected nodes within a clique is treated exactly, while the contribution of infected nodes in external cliques is approximated (i.e. the terms involving \( \rho \)). We thus call this an approximate master equation system.

The system eventually settles to a stationary state in the limit \( t \to \infty \), and henceforth we assume that the quantities \( s_m, c_{n,i}, r \) and \( \rho \) have reached a fixed point. These variables characterizing the stationary state are obtained by solving the following self-consistent expressions

\[ s_m = \frac{1}{1 + mr}, \quad (3a) \]

\[ (i+1)c_{n+1,i} = [i + (n-i)[\beta + \rho]]c_{n,i} + \sum_j \delta_{ij} \rho \quad (s_0 = 1 - \sum c_{n,i}), \quad (3b) \]

which are derived from Eq. (1), and where \( r \) and \( \rho \) are still obtained from Eq. (2). Note that it will be useful to rewrite Eq. (3b) more explicitly as

\[ c_{n,i} \equiv c_{n,0} \frac{n!}{(n-i)!} \prod_{j=0}^{i-1} [\beta j + \rho] \quad \forall i \in \{1, \ldots, n\}, \quad (4) \]

with \( c_{n,0} = 1 - \sum_{i=1}^n c_{n,i} \).

### A. Epidemic threshold

For the SIS dynamics, there exists a critical value \( \beta_c \) for the transmission rate, called the epidemic threshold. For \( \beta < \beta_c \), the absorbing-state—where all nodes are susceptible—is attractive for all initial conditions. For \( \beta > \beta_c \), the absorbing-state becomes unstable and there exists a non-trivial stationary state.

To obtain an expression for \( \beta_c \), let us use \( \rho \) as a reference point and redefine \( r(\rho), s_m(\rho) \) and \( c_{n,i}(\rho) \). We then define the right-hand side of Eq. (2b) as \( F(\rho) \) and a positive solution \( \rho = F(\rho) \) exists if

\[ \frac{dF}{d\rho}\bigg|_{\rho=0} > 1. \]

At the epidemic threshold \( \beta_c \), the above derivative is exactly 1, resulting in \( \rho \to 0 \), \( r(\rho) \to 0 \), \( s_m(\rho) \to 1 \) and \( c_{n,i}(\rho) \to \delta_{0,0} \), where \( \delta_{ij} \) is the Kronecker delta.

It will prove useful to expand \( c_{n,i} \) near the epidemic threshold as \( c_{n,i}(\rho) = h_{n,i,0} + O(\rho^2) \). From Eq. (4), we obtain

\[ h_{n,i} \equiv \frac{dc_{n,i}}{d\rho}\bigg|_{\rho=0} = n!g^{i-1}(i-1)! \quad (n - i)! \quad \forall i \in \{1, \ldots, n\}, \]

and by definition \( h_{n,0} \equiv -\sum_{i=1}^n h_{n,i} \).

For all \( n \), we encode each sequence \( \{h_{n,i}\}_{i=0}^n \) in the generating function

\[ H_n(x; \beta) = \sum_i h_{n,i} x^i, \]

\[ = h_{n,0} + \sum_{i=1}^n \frac{n!}{(n-i)!} (\beta x)^i (i-1)! \]

\[ = h_{n,0} + \frac{1}{\beta} \int_0^{\infty} \frac{n!}{(n-i)!} (\beta u)^i (i-1)! e^{-u} du, \]

\[ = h_{n,0} + \frac{1}{\beta} \int_0^{\infty} [(1 + \beta xu)^n - 1] u^{-1} e^{-u} du. \quad (5) \]

Interestingly, the auxiliary generating function

\[ Q_n(x; \beta) = \frac{H_n(x; \beta) - h_{n,0}}{H_n(1; \beta) - h_{n,0}} = \frac{\int_0^{\infty} [(1 + \beta xu)^n - 1] u^{-1} e^{-u} du}{\int_0^{\infty} [(1 + \beta u)^n - 1] u^{-1} e^{-u} du}, \]

can be interpreted as the probability generating function for the quasi-stationary distribution (only for \( i > 0 \)) for the number of infected nodes in a clique of size \( n \), under the influence of a weak (vanishing) external field.

The introduction of these generating functions allows us to write

\[ \frac{dF}{d\rho}\bigg|_{\rho=0} = \beta \frac{m(m-1)}{\langle m \rangle \langle n \rangle}, \quad (n-1)H_n'(1; \beta) - H_n'(1; \beta) \]

where we have used standard properties of generating functions in combination with Eqs. (2a) and (2b). We simplify the above equation by noting that

\[ (n-1)H_n'(1; \beta) - H_n'(1; \beta) = n(n-1) \int_0^{\infty} (1 + \beta u)^{n-2} e^{-u} du. \]

The epidemic threshold \( \beta_c \) is thus obtained by solving the following implicit equation for \( \beta \)

\[ \beta \frac{m(m-1)}{\langle m \rangle \langle n \rangle} - (n-1)A_n(\beta) = 1. \]
where

\[ A_n(\beta) \equiv \int_0^\infty (1 + \beta u)^n e^{-u} du . \]  

(7)

Note that \( A_n \) could be rewritten in terms of the upper incomplete gamma function, but this integral representation will be more useful later on.

Although it is not possible to write \( \beta_c \) in closed form, we obtain lower and upper bounds on the epidemic threshold

\[ \beta_c \geq \frac{1}{\Omega(g_m, p_n) + (n_{\text{max}} - 2)} , \]  

(8a)

\[ \beta_c \leq \frac{1}{\Omega(g_m, p_n)} , \]  

(8b)

where we defined the coupling between cliques as

\[ \Omega(g_m, p_n) \equiv \left( \frac{\langle m(m-1) \rangle}{\langle m \rangle} \right) \left( \frac{\langle n(n-1) \rangle}{\langle n \rangle} \right) . \]  

(9)

Indeed, if we were to take a random node within a clique, \( \Omega(g_m, p_n) \) would correspond to its average number of external neighbors. It is therefore a good measure of the interaction of cliques with one another. See Appendix A for further details on the derivation of Eq. (8).

B. Behavior for heterogeneous membership and clique size

Let us consider power-law distributions \( p_n \propto n^{-\gamma_n} \) and \( g_m \propto m^{-\gamma_m} \) with large cut-offs \( n_{\text{max}} \gg 1 \) and \( m_{\text{max}} \gg 1 \). We set \( \gamma_n, \gamma_m > 2 \) so that \( \langle n \rangle \) and \( \langle m \rangle \) remain bounded. Figure 2(a) and Fig. 2(d), we note that all \( s_m \) decrease faster in the former case as the ratio \( \beta/\beta_c \) increases. From Eq. (3a), this is explained by a faster increase of the mean field \( r \), resulting directly from a stronger coupling between cliques—i.e. a larger value of \( \Omega(g_m, p_n) \).

The difference between Fig. 2(b) and 2(e) is more striking. While the fraction of infected nodes within cliques \( I_c \) does not vary much with \( n \) in Fig. 2(b)—the coupling \( \Omega(g_m, p_n) \) is strong—we observe a sequential activation of the cliques for the weakly coupled system in Fig. 2(e). Figures 2(c) and 2(f) provide an even clearer illustration for a fixed \( \beta \). When the coupling is strong, all distributions \( c_{n,i} \) are concentrated around roughly the same fraction of infected nodes within the cliques. Weak coupling yields a more diverse scenario where smaller cliques have very few infected nodes while the prevalence in larger cliques can be very high. We qualify the latter as active cliques. For cliques of moderate size (e.g. \( n = 50 \)), some \( c_{n,i} \) are highly dispersed, akin to a system near a critical point. This is a telling illustration of why dynamics on networks with higher order structures require approximate master equation approaches: Structures can have heterogeneous state distributions, and a cruder approximation (e.g., models averaging \( i/n \) for all cliques of a given size or mean-field approximations) is likely to miss many rich features of the dynamics.

The scenario presented by Figs. 2(e) and 2(f) is typical of a smeared phase transition. Instead of clean critical point driven by a collective ordering, subparts of the system self-activate independently from the rest, as shown by the local order parameters \( I_n \). This behavior has an intuitive explanation. Since \( p_n \propto n^{-\gamma_n} \), a small proportion of the cliques are very large, albeit of finite size. Near \( \beta_c \), the largest cliques are able to self-sustain an endemic state by themselves, but since the coupling is weak, the contagion does not spread through the rest of the network. As \( \beta \) increases beyond \( \beta_c \), more cliques are able to self-sustain a local outbreak, until a point where the epidemics delocalizes and invades the whole network. This analytical description is in line with the work of Ref. [26], where numerical evidence for Griffiths phases was found in a similar setting.

To predict the emergence of this phenomenon, we need to have some better intuition on the behavior of \( I_n \) near \( \beta_c \). Since \( \rho \to 0 \) near the critical point, we write \( I_n = H'_n(1) \rho/n + O(\rho^2) \). Performing a saddle-point approximation for \( H'_n(1) \), we obtain the following asymptotic behavior for large clique sizes \( n \)

\[ H'_n(1) \sim \begin{cases} \frac{n}{1 - \beta n} & \text{if } \beta < (n - 1)^{-1} \\ n^{3/2} (\beta n)^{n} e^{-n+1/\beta} & \text{if } \beta \geq (n - 1)^{-1} \end{cases} \]  

(10)

For \( \beta = a(n - 1)^{-1} \) where \( a > 1 \) is a constant independent from \( n \), this implies that \( I_n = O(n^{1/2} e^m) \) with \( b > 0 \). A more formal proof could be made following an argument similar to the one used in Appendix A. This means that near the epidemic threshold (i.e. \( \beta = \beta_c + \epsilon \) with \( \epsilon \ll 1 \)), we expect the epidemic to be localized within any cliques of size \( n \) for which \( \beta > (n - 1)^{-1} \). More formally, we say that the epidemic is localized near the epidemic threshold when \( I_{\text{max}}/I_2 = O(n^{1/2} e^m) \), and we then expect a smeared phase transition, such as the one presented in Fig. 2(e). Conversely, if \( I_{\text{max}}/I_2 = O(1) \) near \( \beta_c \), then we say that the epidemic is delocalized, and we expect a phase transition similar to the one shown on Fig. 2(b).

III. MESOSCOPIC LOCALIZATION

In this section, we fully characterize the emergence of mesoscopic localization, where the epidemic is localized only within the largest cliques near \( \beta_c \) and where memberships and clique sizes are distributed according to power-law distributions. In Sec. III A we derive general asymptotic expressions to distinguish the localization regimes, forming a partition of the \((\gamma_m, \gamma_n)\) space. We then investigate in Sec. III B the effect of finite cut-offs on the localization regimes, and how our results relate to earlier works using the inverse participation ratio.

A. Asymptotic localization regimes

Let us assume that both \( n_{\text{max}} \to \infty \) and \( m_{\text{max}} \to \infty \). As it will be shown, the relation between the cut-offs \( n_{\text{max}} \) and \( m_{\text{max}} \)}
influences the localization regimes. Henceforth, let us assume a general asymptotic relationship of the form $m_{\text{max}} \sim n_{\text{max}}^\alpha$, where the exponent $\alpha \geq 0$ encodes how both limits $n_{\text{max}} \to \infty$ and $m_{\text{max}} \to \infty$ are taken.

To gain some insights on what this relationship between the cut-offs represents, let us assume for a moment that we have a finite-size network with $N$ nodes and $N' = (m)N/(n) \sim N$ cliques. We could impose cut-offs that are agnostic to the underlying distribution $g_m$ and $p_n$, for instance $m_{\text{max}} \sim N^{1/2}$ and $n_{\text{max}} \sim N^{1/2}$. This would correspond to $\alpha = 1$. Another option, borrowed from extreme value theory, would be to use the natural cut-offs of the two power-law distributions, $m_{\text{max}} \sim N^{1/(\gamma_m-1)}$ and $n_{\text{max}} \sim N^{1/(\gamma_n-1)}$ [36][37], yielding $\alpha = (\gamma_n - 1)/(\gamma_m - 1)$. Finally, fixing one of the two cut-offs while letting the other go to infinity would correspond to the limit cases $\alpha \to 0$ or $\alpha \to \infty$.

We now turn to the extraction of the asymptotic behavior of the epidemic threshold in the limit $n_{\text{max}} \to \infty$ for different combinations of $\gamma_n$ and $\gamma_m$—this will inform us on the type of phase transition, i.e. a localized or a delocalized one. First, we obtain a tighter upper-bound on $\beta_c$ in the limit $n_{\text{max}} \to \infty$ for power-law clique size distributions $p_n \propto n^{-\gamma_n}$,

$$\beta_c \leq \min \left\{ \frac{1}{\Omega(g_m, p_n)}, \frac{1}{n_{\text{max}}^{-2}} \right\}. \quad (11)$$

Details are provided in Appendix [A] but the general idea is to combine Eq. (8a) with another bound found by forbidding $A_{\text{max}}$ to grow exponentially with $n_{\text{max}}$.

The lower bound of Eq. (8a) and the upper bound of Eq. (11) tightly constrain the asymptotic behavior of $\beta_c$, which we write as

$$\beta_c^{-1} \sim \Omega(g_m, p_n) + n_{\text{max}}.$$

The first factor in Eq. (9) has the following behavior

$$\frac{\langle m(m-1) \rangle}{\langle m \rangle} \sim \begin{cases} \frac{\alpha^{(3-\gamma_m)}}{n_{\text{max}}} & \text{if } \gamma_m < 3, \\ \alpha \ln n_{\text{max}} & \text{if } \gamma_m = 3, \\ 1 & \text{if } \gamma_m > 3, \end{cases} \quad (12)$$

FIG. 2. Comparison of the stationary state near the epidemic threshold in the delocalized and mesoscopic localization regimes. Stationary state solutions were obtained from Eqs. (3)(a) and (4) for heterogeneous membership and clique size distributions of the form $g_m \propto m^{-\gamma_m}$ and $p_n \propto n^{-\gamma_n}$. We used minimal values $m_{\text{min}} = n_{\text{min}} = 2$ and finite cut-offs $m_{\text{max}} = n_{\text{max}} = 100$. The epidemic threshold $\beta_c$ is the solution to Eq. (5). (a)-(d) Stationary fraction of susceptible nodes with a given membership $m$ to use the other option, borrowed from extreme value theory, would be finite-size network with $N$ nodes and $N' = (m)N/(n) \sim N$ cliques. We could impose cut-offs that are agnostic to the underlying distribution $g_m$ and $p_n$, for instance $m_{\text{max}} \sim N^{1/2}$ and $n_{\text{max}} \sim N^{1/2}$. This would correspond to $\alpha = 1$. Another option, borrowed from extreme value theory, would be to use the natural cut-offs of the two power-law distributions, $m_{\text{max}} \sim N^{1/(\gamma_m-1)}$ and $n_{\text{max}} \sim N^{1/(\gamma_n-1)}$ [36][37], yielding $\alpha = (\gamma_n - 1)/(\gamma_m - 1)$. Finally, fixing one of the two cut-offs while letting the other go to infinity would correspond to the limit cases $\alpha \to 0$ or $\alpha \to \infty$.

We now turn to the extraction of the asymptotic behavior of the epidemic threshold in the limit $n_{\text{max}} \to \infty$ for different combinations of $\gamma_n$ and $\gamma_m$—this will inform us on the type of phase transition, i.e. a localized or a delocalized one. First, we obtain a tighter upper-bound on $\beta_c$ in the limit $n_{\text{max}} \to \infty$ for power-law clique size distributions $p_n \propto n^{-\gamma_n}$,

$$\beta_c \leq \min \left\{ \frac{1}{\Omega(g_m, p_n)}, \frac{1}{n_{\text{max}}^{-2}} \right\}. \quad (11)$$

Details are provided in Appendix [A] but the general idea is to combine Eq. (8a) with another bound found by forbidding $A_{\text{max}}$ to grow exponentially with $n_{\text{max}}$.

The lower bound of Eq. (8a) and the upper bound of Eq. (11) tightly constrain the asymptotic behavior of $\beta_c$, which we write as

$$\beta_c^{-1} \sim \Omega(g_m, p_n) + n_{\text{max}}.$$

The first factor in Eq. (9) has the following behavior

$$\frac{\langle m(m-1) \rangle}{\langle m \rangle} \sim \begin{cases} \frac{\alpha^{(3-\gamma_m)}}{n_{\text{max}}} & \text{if } \gamma_m < 3, \\ \alpha \ln n_{\text{max}} & \text{if } \gamma_m = 3, \\ 1 & \text{if } \gamma_m > 3, \end{cases} \quad (12)$$
FIG. 3. Asymptotic localization regimes for heterogeneous membership and clique size distributions of the form $g_m \sim n^{\gamma_m}$ and $p_n \sim n^{\gamma_n}$. We assume different cut-off relationships of the form $m_{\text{max}} \sim n^{\gamma_n}$. In the pale green region, the epidemic is localized near the epidemic threshold $\beta_c$, while it is delocalized in the darker blue regions. The boundary separating the two regions is inferred from Eqs (14a), (14b) and (14c).

and the second one has a similar form

$$\frac{(n(n-1))}{n} \sim \begin{cases} n_{\text{max}}^{3-\gamma_m} & \text{if } \gamma_n < 3, \\ \ln n_{\text{max}} & \text{if } \gamma_n = 3, \\ 1 & \text{if } \gamma_n > 3. \end{cases}$$

Combining Eqs. (12) and (12) for different $\gamma_n$ and $\gamma_m$ leads to different scalings for $\Omega(g_m, p_n)$ and, as a result, different scalings for $\beta_c$. We distinguish three asymptotic behaviors:

1. $\Omega(g_m, p_n)n_{\text{max}}^{-1} \rightarrow \infty \Rightarrow \beta_c n_{\text{max}} \rightarrow 0$,
2. $\Omega(g_m, p_n)n_{\text{max}}^{-1} = O(1) \Rightarrow \beta_c n_{\text{max}} \rightarrow q < 1$,
3. $\Omega(g_m, p_n)n_{\text{max}}^{-1} \rightarrow 0 \Rightarrow \beta_c n_{\text{max}} \rightarrow 1$.

If $\gamma_n \geq 3$, we necessarily have $\beta_c n_{\text{max}} \rightarrow 1$. Otherwise,

- If $\gamma_n < 3$, then
  $$\beta_c n_{\text{max}} \rightarrow \begin{cases} 0 & \text{if } 3 - \gamma_n + \alpha(3 - \gamma_m) > 1, \\ q & \text{if } 3 - \gamma_n + \alpha(3 - \gamma_m) = 1, \\ 1 & \text{if } 3 - \gamma_n + \alpha(3 - \gamma_m) < 1. \end{cases}$$ (14a)

- If $\gamma_n = 3$, then
  $$\beta_c n_{\text{max}} \rightarrow \begin{cases} 0 & \text{if } \alpha(3 - \gamma_m) \geq 1, \\ 1 & \text{if } \alpha(3 - \gamma_m) < 1. \end{cases}$$ (14b)

- If $\gamma_n > 3$, then
  $$\beta_c n_{\text{max}} \rightarrow \begin{cases} 0 & \text{if } \alpha(3 - \gamma_m) > 1, \\ q & \text{if } \alpha(3 - \gamma_m) = 1, \\ 1 & \text{if } \alpha(3 - \gamma_m) < 1. \end{cases}$$ (14c)

Let us note that the category $\beta_c n_{\text{max}} \rightarrow q < 1$ never fills an area in the $(\gamma_n, \gamma_m)$ space—it is only a limit case—whereas the other two other scenarios split the space into two separate regions. In the region where $\beta_c n_{\text{max}} \rightarrow 0$, right above the epidemic threshold we have $I_{\text{max}}/I_2 = O(1)$, hence the epidemic is delocalized. In the region where $\beta_c n_{\text{max}} \rightarrow 1$, we have instead $I_{\text{max}}/I_2 = O(n_{\text{max}}^{1/2} e^{\beta_c n_{\text{max}}})$ and the epidemic is localized.

FIG. 4. Impact of finite cut-offs on the boundary separating the localized and delocalized regimes for power-law membership and clique size distributions. We used minimal values $m_{\text{max}} = n_{\text{max}} = 2$ and same cut-offs $m_{\text{max}} = n_{\text{max}}$, hence $\alpha = 1$, and different values of $n_{\text{max}}$. Finite cut-offs boundaries are obtained by imposing $\beta_c = n_{\text{max}}^{-1}$ and solving Eq. (6) for different pairs $(\gamma_n, \gamma_m)$. The asymptotic case $n_{\text{max}} \rightarrow \infty$ is obtained from Eqs (14a).

Equations (14a), (14b) and (14c) can then be used to distinguish the region where we expect mesoscopic localization, as illustrated in Fig. 3 for different values of $\alpha$. One striking observation is the ubiquity of mesoscopic localization: for a large portion of the parameter space, we expect a disease to be localized around the largest cliques. It is worth to recall that the average degree of a node is proportional to $(n(n-1))$, hence sparse networks correspond only to the upper portion ($\gamma_n > 3$) of the phase diagrams in Fig. 3.

B. Finite cut-offs and mesoscopic inverse participation ratio

The results of Sec. [III A] were obtained in the asymptotic limit $n_{\text{max}} \rightarrow \infty$. However, whether it is due to some inherent properties or simply the finite size of the system, we expect fi-
nite cut-offs $n_{\text{max}}$ and $m_{\text{max}}$, often smoothed by an exponential decay—we keep hard cut-offs for mathematical convenience.

A finite value for $n_{\text{max}}$ relaxes the conditions (14a), (14b) and (14c). For a pair $(\gamma_m, \gamma_n)$ in the asymptotically localized regime, it is possible to have either $\beta_c \approx n_{\text{max}}^{-1}$. To stay coherent with our definition for a localized epidemic, we must have $\beta_c > n_{\text{max}}^{-1}$. Therefore, a finite value for $n_{\text{max}}$ effectively changes the boundary defined by Eqs. (14a), (14b) and (14c) separating the two regimes. This new curve can be obtained by solving Eq. (6) for pairs $(\gamma_m, \gamma_n)$ that satisfy $\beta_c \equiv n_{\text{max}}^{-1}$. In Fig. 4, we illustrate the boundary separating the delocalized and localized regimes for increasing values of $n_{\text{max}}$, slowly converging on the asymptotic conditions. The region of the parameter space corresponding to mesoscopic localization is shrunk compared to the asymptotic limit, but it still fills most of the parameter space corresponding to sparse networks.

Another consequence of finite cut-offs is to blur the line between localized and a delocalized epidemic. Taking pairs $(\gamma_m, \gamma_n)$ closer to the finite-size boundary, we show how this affects the clique prevalence in Fig. 5(a) and Fig. 5(b), with $\beta_c < n_{\text{max}}^{-1}$ and $\beta_c > n_{\text{max}}^{-1}$, respectively. Near $\beta_c$, we still associate Fig. 5(a) and 5(b) with a delocalized and localized outbreak respectively, but the difference is less striking compared to Fig. 4(a) and 4(b). Therefore, even though the dichotomy is sharp and clear in the asymptotic limit $n_{\text{max}} \to \infty$, we need to keep in mind that for realistic systems, localization lives on a spectrum, one that we should try to quantify.

At the level of nodes, an epidemic is considered localized if the contagion is mostly present within a subset of nodes $\mathcal{L} \subset \mathcal{V}$, referred to as the localization set, and $\mathcal{V} = \{1, \ldots, N\}$ is the set of all nodes. An important result from quenched mean-field theory is that the marginal probability for each node $j$ of being infected near $\beta_c$ is proportional to $v_j$, where $\{v_j\}_{j \in \mathcal{V}}$ are the elements of the principal eigenvector (PEV) of the adjacency matrix. Epidemic localization can thus be mapped onto eigenvector localization [21, 38, 41]. With a normalized eigenvector satisfying $\sum_j v_j^2 = 1$, a completely delocalized epidemic at the level of nodes implies $v_j \sim N^{-1/2} \forall j \in \mathcal{V}$, while a purely localized one corresponds to $v_j \sim |\mathcal{L}|^{-1/2} \forall j \in \mathcal{L}$ and $v_j \sim 0 \forall j \notin \mathcal{L}$. A standard scalar to quantify the localization is the inverse participation ratio $Y_{\mathcal{L}}(N)$; here we use a rescaled version of this inverse participation ratio

$$Y_{\mathcal{L}}(N) \equiv N \sum_{j=1}^{N} v_j^2 .$$

For a delocalized eigenvector, $Y_{\mathcal{L}}(N) \sim 1$, while for localization set of size $|\mathcal{L}| \sim N^\nu$, then $Y_{\mathcal{L}}(N) \sim N^{1-\nu}$. Consequently, $Y_{\mathcal{L}}^{-1}$ is an effective measure for the fraction of nodes belonging to the localization set.

At the mesoscopic level, we consider an epidemic localized if the contagion is mostly present within a subset of the higher-order structures, more specifically in this paper—within cliques of a certain size $n$. The difference is subtle, but it is important: if we observe a delocalized epidemic at the mesoscopic level, it could still be localized at the level of nodes. To quantify mesoscopic localization, we use an inverse participation ratio as well

$$\tilde{Y}_{4}(p_n) = \frac{\sum_{n} |v_n|^2}{(\sum_{n} |v_n|^2)^2} = \begin{cases} 1 & \text{if } I_n \gg 1 \forall n , \\ |p_n^{-1}| & \text{if } I_n \ll \delta_{n,n'} . \end{cases}$$

As a result, $Y_{4}^{-1}$ is an effective measure for the fraction of cliques participating to the epidemic. Interestingly, Eq. (16) can be obtained with our analytical formalism, using $I_n$ evaluated at the epidemic threshold $\beta_c$, or through the connection with quenched mean-field theory. In the latter case, one extracts the PEV of a network with cliques, then compute

$$I_n \propto \left| C_n \right| \sum_{S \subseteq C_n} \sum_{j \in S} v_j ,$$

where $S$ is the set of nodes belonging to a clique and $C_n$ is the set of cliques of size $n$. The clique distribution is on its hand $p_n \sim |C_n|$. Note that this measure relies on an explicit knowledge of $C_n$, which is automatically given for synthetic networks (see Appendix B), or could be extracted using a clique decomposition. The concept obviously generalizes to various kinds of mesoscopic structures, not just cliques.

In Fig. 5(c), we illustrate the behavior of $Y_{\mathcal{L}}(p_n)$ as a function of $(\gamma_m, \gamma_n)$, obtained with our analytical formalism. As expected, the inverse participation changes drastically near the boundary separating the delocalized and localized regimes for finite cut-offs. The change would become sharper and sharper as we let $n_{\text{max}} \to \infty$, and the position of the boundary would move closer to the asymptotic limit, as in Fig. 4. This inverse participation ratio is therefore a good measure for mesoscopic localization, and could be used to get insights on how the epidemic changes from a localized to a delocalized phase as we increase $\beta$ over $\beta_c$.

In Fig. 6(a), we compare the finite-size scaling of the inverse participation ratios for nodes and cliques, obtained by generating synthetic networks in the delocalized and localized regime and extracting their PEV. Although our analytical formalism effectively describes cliques of a sub-extensive size, this is not a necessary condition to observe mesoscopic localization. We have therefore relaxed this assumption to generate the synthetic networks: we used cut-offs that scale with the number of nodes $m_{\text{max}} = N^{1/(\gamma_{n-1})}$ and $n_{\text{max}} = N^{1/(\gamma_{n-1})}$. These are more appropriate for the finite-size scaling analysis.

In Fig. 6(a), we see that the inverse participation ratio for nodes $Y_{\mathcal{L}}$ increases in both the delocalized and the mesoscopic localization regime. It scales similarly to the inverse fraction of the nodes belonging to the maximal $K$–core, in agreement with previous works on the subject [20, 21, 40]. The localization set can thus be associated with the innermost core in both cases, and despite a different scaling law, there is no clear sign of a change of regime between the two curves. Figure 6(b) tells us another story: the inverse participation ratio for cliques $\tilde{Y}_{4}$ converges to 1 in the delocalized regime, but scales as a power law in the mesoscopic localization regime, clearly indicating a transition of regime.

Figure 6 strongly advocates for a change of perspective if we want to detect potentially hidden localized phase at the mesoscopic level: we need to focus on these higher-order
substructures explicitly and find better ways to characterize their impact on the dynamics. If we focus our attention at the node level, Fig. 5(a) tells us that an epidemic localized at the mesoscopic level is no different from a delocalized one—the contagion is mostly present within the innermost core in both cases. However, the nature of this core and of the outer shells are much different, as can be inferred from \( \tilde{Y}_4 \) in Fig. 5(b). In the localized regime, the innermost core is composed mostly of the largest cliques, while cliques of all sizes compose the core in the delocalized regime. Let us recall that a bias toward larger and denser structures has dramatic consequences on the dynamics, leading to a smeared phase transition instead of a clean one.

**IV. DISCUSSION**

One of the key factors behind the success of network science to study contagions, from infectious diseases to the spread of information, is that it provides a mathematical framework to go beyond the assumption of a homogeneous population [7]. Contagions are rarely driven by the average individual, mostly because some individuals are simply more connected than others but also potentially more central. Beyond the fact that they drive the dynamics of contagions, these key actors are also critical to their control. It allows, on the one hand, to mathematize methods of targeted immunization and interventions [42, 43]. Which individuals should be immunized or removed from the network to minimize the spread on an infectious disease? It also allows, on the other hand, to identify influential spreaders [44]. Which individuals should seed a contagion in order to maximize its spread? These different ideas all revolve around the idea of control theory for contagions, but also all depend on a good theoretical understanding of what type of structure matters for contagions.

In practice, however, social networks are not randomly mixed but contain a high level of structure determined by social groups, workplaces, schools and events; such that key actors can be places and groups rather than the individuals themselves. Thankfully, multiple new approaches to handle higher-order interactions have been proposed in recent years. In the thermodynamic limit, the clique-based networks used in this paper can equivalently be represented using ideas of topological simplexes from topology [45], hypergraphs [46, 47], or projections of bipartite networks [17, 34]. Up to the right level of mean-field approximation, these are all equivalent. However, their dynamics at the mesoscopic level can be very het-

---

**FIG. 5.** Comparison of the stationary state and the level of localization near the epidemic threshold for networks closer to the finite cut-offs boundary. We used power-law membership and clique size distributions with minimal values \( m_{\text{min}} = n_{\text{min}} = 2 \) and cut-offs \( m_{\text{max}} = n_{\text{max}} = 100 \). (a)-(b) Clique prevalence (solid lines) and global prevalence (dashed line) as a function of the transmission rate. (a) \( \gamma_m = \gamma_n = 2.6 \), yielding \( \beta_c < n_{\text{max}}^{-1} \). (b) \( \gamma_m = \gamma_n = 3.1 \), yielding \( \beta_c > n_{\text{max}}^{-1} \). (c) Quantification of the mesoscopic localization phenomenon using the inverse participation ratio defined at Eq. (16). The solid line corresponds to the boundary between the localized and delocalized regimes, obtained by imposing \( \beta_c = n_{\text{max}}^{-1} \) and solving Eq. (6) for different pairs \( (\gamma_m, \gamma_n) \).

**FIG. 6.** Mesoscopic localization is imperceptible using the standard inverse participation ratio on nodes. We performed a finite-size scaling of the inverse participation ratio in the delocalized and mesoscopic localization regime, for nodes and cliques. We generated multigraphs of various sizes with different power-law membership and clique size distributions [see Appendix B]. In the delocalized regime, we used \( \gamma_m = 2.3 \) and \( \gamma_n = 3.5 \); in the localized regime we used \( \gamma_m = \gamma_n = 3.5 \). We used minimal values \( m_{\text{min}} = n_{\text{min}} = 2 \) and natural cut-offs \( m_{\text{max}} = N^{1/(\gamma_m-1)} \) and \( n_{\text{max}} = N^{1/(\gamma_n-1)} \). (a) The solid lines represent the average inverse participation ratio \( \tilde{Y}_4 \) for nodes obtained with Eq. (15). The dashed (dotted) line is the average inverse of the fraction of nodes associated to the maximal \( K \)-core in the delocalized (localized) regime. (b) The solid lines represent the average inverse participation ratio \( \tilde{Y}_4 \) for cliques. We extracted \( I_n \) from the PEV, then used Eq. (16). The shaded regions in both panels correspond to one standard deviation.
erogeneous, as in Fig. 2(f), since substructures can take many more different states than individuals who are usually only susceptible or infected. We therefore have to avoid coarse-graining the mathematical description too much in order to embrace this heterogeneity, as we did with this clique-based approximate master equation framework.

Using this approach, we observed and analyzed a phenomenon of mesoscopic localization where contagions can concentrate around key substructures that are large enough to allow a local, self-sustained outbreak with the help of some weak external coupling. Interestingly, while there is little empirical evidence for localization of real contagions around hubs in a contact network, there are well-known cases of dynamics resembling mesoscopic localization. For example, bacterial infections in hospitals (e.g., C. difficile [48]) are already a well-documented example of mesoscopically localized contagions but are simply never studied analytically as such.

In this mesoscopic localization phase, influential structures are naturally found to be the larger ones around which a contagion can localize. Intervention or control operating at a structural level (i.e., on groups rather than on individuals) should therefore focus around these influential structures. The large toolbox developed for targeted immunization [49] and identification of influential spreaders [18] could now be leveraged, at the mesoscopic level, to better understand and control contagions on networks capable of this mesoscopic localization. In Ref. [33], we investigate the impact of removing cliques as a model of school closures and event cancellations. We find that close to their epidemic threshold, de-localized dynamics are characterized by a linear relationship between outbreak size and the scale of our intervention measured in number of edges removed; conversely, localized dynamics show a non-linear relationship such that there are increasing returns on edges removed; conversely, localized dynamics show a non-linear relationship such that there are increasing returns on edges removed; conversely, localized dynamics show a non-linear relationship such that there are increasing returns on edges removed. Our results suggest that structural interventions are potentially more efficient than individual interventions to control localized dynamics.

More broadly, higher-order structures were found to be important for a wide range of dynamics, from competitive dynamics [50] to social contagion [45]. Several of these studies highlight non-trivial effects of higher-order structure on dynamics using numerical tools or very coarse-grained analytical tools. These tools, by ignoring the heterogeneous states of network substructures, limit the type of questions and behaviors that can be answered and analyzed. We therefore think that master equation descriptions could provide deep insights into the mechanisms of these dynamics and their interplay with higher-order structure. For instance, we conjecture that mesoscopic localization is even more ubiquitous in systems with social reinforcement mechanisms [51], and that its impacts on the global state of the dynamics are even more dramatic.

In fact, there are several avenues now open to broaden the applicability of our approach. As a first step, our future works will focus on improving the heterogeneous mean-field coupling between master equations. We could, for example, include more information in our description of node states in order to capture dynamical correlations between the state of nodes and the state of the structures in which they are found; or include correlations between the memberships of nodes and the sizes of structures; or allow structures with more complex inner contact patterns. Altogether, we hope that our work on mesoscopic localization should provide a useful roadmap to improve the study of dynamics on networks with higher-order structures.

ACKNOWLEDGMENTS

The authors acknowledge Calcul Québec for computing facilities. This work was supported by the National Institutes of Health 1P20 GM125498-01 Centers of Biomedical Research Excellence Award (L.H.-D.), the Fonds de recherche du Québec – Nature et technologies (V.T., G.S.), the Natural Sciences and Engineering Research Council of Canada (G.S., V.T., A.A., L.J.D.), and the Sentinelle Nord program of Université Laval, funded by the Canada First Research Excellence Fund (G.S., V.T., A.A., L.J.D.).

Appendix A: Bounds on the epidemic threshold

Let us first bound \( \beta_{\text{max}} \) for any \( n_{\text{max}} \) by bounding \( A_n \) for all \( n \). First, since \( \beta u \geq 0 \), then

\[
A_n(\beta) \geq \int_0^\infty e^{-u} du = 1. \tag{A1}
\]

Secondly, we rewrite

\[
A_n(\beta) = \int_0^\infty e^{\phi(u;\beta)} du,
\]

where \( \phi(u;\beta) = (n - 2) \ln(1 + \beta u) - u \). Since \( \ln(1 + x) \leq x \), then \( \phi(u;\beta) \leq (n - 2) u - u \), which implies

\[
A_n(\beta) \leq \begin{cases} 
\infty & \text{if } \beta(n - 2) \geq 1, \\
1 & \text{if } \beta(n - 2) < 1.
\end{cases}
\]

We relax the conditions by replacing \( n \) by \( n_{\text{max}} \) everywhere on the right-hand side

\[
A_n(\beta) \leq \begin{cases} 
\infty & \text{if } \beta(n_{\text{max}} - 2) \geq 1, \\
\frac{1}{1 - \beta(n_{\text{max}} - 2)} & \text{if } \beta(n_{\text{max}} - 2) < 1.
\end{cases}
\tag{A2}
\]

By inserting Eqs. (A1) and (A2) in Eq. (6) and solving for \( \beta \), we find the bounds of Eq. (8). Only the second case of Eq. (A2) leads to a coherent bound for \( \beta_c \).

The upper bound on the epidemic threshold is not very tight, but we can do better if we assume \( p_n \propto n^{-\alpha} \) and the limit \( n_{\text{max}} \to \infty \). It follows that the epidemic threshold must respect \( \beta_c \leq (n_{\text{max}} - 2)^{-1} \). Let us make a proof by contradiction: we start with the premise that \( \beta = o(n_{\text{max}} - 2)^{-1} \) for some arbitrary constant \( \alpha > 1 \). We know that \( \ln(1 + x) \geq x(1 - x) \) for all \( x \geq 0 \), hence

\[
\phi(u;\beta) \geq -\beta^2(n - 2)u^2 + [\beta(n - 2) - 1]u.
\]
Making the change of variable $y = \beta \sqrt{n - 2u}$ and defining $d \equiv [\beta(n - 2) - 1]/(2\beta \sqrt{n - 2})$, we arrive at

$$A_n(\beta) \geq \frac{e^{\frac{d^2}{\beta \sqrt{n - 2}}}}{\sqrt{\pi \beta}} \int_0^\infty e^{-(y-d)^2} dy \quad \forall n > 2,$$

$$\geq \frac{\sqrt{\pi} e^{d^2}}{2\beta \sqrt{n - 2}}.$$

Let us focus on $n = n_{\text{max}}$. In this case, using our premise for $\beta$, we have

$$d^2 = \frac{(a-1)^2(n_{\text{max}} - 2)}{4a^2} \equiv b(n_{\text{max}} - 2),$$

where $b > 0$. Therefore, there always exists a constant $B_1 > 0$ independent from $n_{\text{max}}$ and $a$ such that

$$A_{n_{\text{max}}}(\beta) \geq \frac{1}{a} \left(\frac{1}{2} \sqrt{\pi(n_{\text{max}} - 2)e^{-2b}}\right)^{e^{2b_{n_{\text{max}}}}} \geq \frac{B_1}{a} e^{b_{n_{\text{max}}}},$$

This provides a lower bound for the following term

$$\langle n(n - 1)A_n(\beta) \rangle \geq p_{n_{\text{max}}} n_{\text{max}}(n_{\text{max}} - 1)A_{n_{\text{max}}}(\beta),$$

$$\geq \frac{B_2}{a} n_{\text{max}}^{2-\gamma_e} e^{b_{n_{\text{max}}}},$$

where we assumed $p_n \propto n^{\gamma_e}$ with $\gamma_e < \infty$. For some constant $B_2 < \infty$, inserting this and our premise on $\beta$ in Eq. (6), we obtain an expression of the form

$$n_{\text{max}}^{1-\gamma_e} \leq B_3 e^{-b_{n_{\text{max}}}} . \quad (A3)$$

For some constant $B_3 < \infty$, Equation (A3) is clearly not respected in the limit $n_{\text{max}} \to \infty$, hence completing the proof by contradiction.

Note that a solution $\beta_c > (n_{\text{max}} - 2)^{-1}$ is not ruled out if $p_n$ decrease exponentially for large $n$.

### Appendix B: Generation of networks with cliques

We generated multigraphs using a stub matching process. First, each node $j \in \mathcal{V}$ is assigned a membership $m$ drawn from $g_m$, resulting in a membership sequence $m = (m_1, m_2, \ldots, m_N)$. Then, we create a clique size sequence of length $N'$, $n = (n_1, n_2, \ldots, n_N)$, by drawing sizes $n_k$ according to $p_n$. However, we must additionally constrain the sequence such that the number of membership stubs and the number of clique stubs (available spot for the nodes) are the same

$$\sum_{j=1}^N m_j = \sum_{k=1}^{N'} n_k . \quad (B1)$$

In practice, if the right-hand side of Eq. (B1) is smaller than the left-hand side, we add another clique with size $n$ drawn from $p_n$. If it is bigger, we remove a clique uniformly at random. We repeat this process until the number of stubs is equal on both sides. Note that $N'$ is therefore not fixed, but it is expected that $N' \sim (m)N/(n)$ since both $(m)$ and $(n)$ are bounded.

Once we have the membership and clique size sequences, we match the stubs uniformly at random—an edge is added between each pair of nodes belonging to a same clique. This effectively creates loopy multigraphs, but the loops and multi-edges represent a vanishing fraction of the total number of edges for $N \to \infty$; we do not remove them since they have a marginal impact on the dynamics.

[1] R. M. Anderson, B. Anderson, and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, 1992).
[2] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, “The legacy of Kermack and McKendrick,” Publ. Newton Inst. 5, 95 (1995).
[3] D. Centola and M. Macy, “Complex contagions and the weakness of long ties,” *Am. J. Sociol.* 113, 702 (2007).
[4] B. Mønsted, P. Sapieżyński, E. Ferrara, and S. Lehmann, “Evidence of complex contagion in information diffusion on social media: An experiment using Twitter bots,” *PLoS One* 12, e0184148 (2017).
[5] S. Lehmann and Y.-. Ahn, *Complex Spreading Phenomena in Social Systems* (Springer, 2018).
[6] L. Hébert-Dufresne and B. M. Althouse, “Complex dynamics of synergistic coinfections on realistically clustered networks,” *Proc. Natl. Acad. Sci. U. S. A.* 112, 10551 (2015).
[7] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, “Epidemic processes in complex networks,” *Rev. Mod. Phys.* 87, 925 (2015) [arXiv:1408.2701].
[8] I. Z. Kiss, J. C. Miller, and P. L. Simon, *Mathematics of Epidemics on Networks: From Exact to Approximate Models*, Vol. 46 (Springer, 2017).
[9] R. Pastor-Satorras and A. Vespignani, “Epidemic spreading in scale-free networks,” *Phys. Rev. Lett.* 86, 3200 (2001).
[10] M. Boguñá and R. Pastor-Satorras, “Epidemic spreading in correlated complex networks,” *Phys. Rev. E* 66, 047104 (2002).
[11] K. T. D. Eames and M. J. Keeling, “Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases,” *Proc. Natl. Acad. Sci. U. S. A.* 99, 13330 (2002).
[12] A. S. Mata, R. S. Ferreira, and S. C. Ferreira, “Heterogeneous pair-approximation for the contact process on complex networks,” *New J. Phys.* 16, 53006 (2014).
[13] V. Marceau, P.-A. Noë, L. Hébert-Dufresne, A. Allard, and L. J. Dubé, “Adaptive networks: Coevolution of disease and topology,” *Phys. Rev. E* 82, 036116 (2010).
[14] J. P. Gleeson, “High-accuracy approximation of binary-state dynamics on networks,” *Phys. Rev. Lett.* 107, 068701 (2011).
[15] J. Lindquist, J. Ma, P. van den Driessche, and F. H. Willeboordse, “Effective degree network disease models,” *J. Math.*
A. V. Goltsev, S. N. Dorogovtsev, J. G. Oliveira, and J. F. F. Mendes, “Localization and spreading of diseases in complex networks,” Phys. Rev. E 68, 026121 (2003).

L. Weng, F. Menczer, and Y.-Y. Ahn, “Predicting successful memes using network and community structure,” in Eighth international AAAI conference on weblogs and social media (2014).

A. V. Goltsev, S. N. Dorogovtsev, J. G. Oliveira, and J. F. F. Mendes, “Community structure in social and biological networks,” Proc. Natl. Acad. Sci. U. S. A. 99, 7821 (2002).

M. E. J. Newman, “Properties of highly clustered networks,” Phys. Rev. E 68, 026121 (2003).

L. Hébert-Dufresne, P.-A. Noël, Vincent Marceau, A. Allard, and L. J. Dubé, “Propagation dynamics on networks featuring complex topologies,” Phys. Rev. E 82, 036115 (2010).

Note that we are still preserving the dynamic correlations between pairs of nodes by tracking each $c_{uv}$.

M. Boguñá, R. Pastor-Satorras, and A. Vespignani, “Cut-offs and finite size effects in scale-free networks,” Eur. Phys. J. B 38, 205 (2004).

M. Catanazzo, M. Boguñá, and R. Pastor-Satorras, “Generation of uncorrelated random scale-free networks,” Phys. Rev. E 71, 027103 (2005).

T. Martin, X. Zhang, and M. E. J. Newman, “Localization and centrality in networks,” Phys. Rev. E 90, 052808 (2014).

R. Pastor-Satorras and C. Castellano, “Distinct types of eigenvector localization in networks,” Sci. Rep. 6, 101038/srep18847 (2016).

C. Castellano and R. Pastor-Satorras, “Relating topological determinants of complex networks to their spectral properties: Structural and dynamical effects,” Phys. Rev. X 7, 41024 (2017).

K. J. Sharkey, “Localization of eigenvector centrality in networks with a cut vertex,” Phys. Rev. E 99, 012315 (2019).

R. Pastor-Satorras and A. Vespignani, “Immunization of complex networks,” Phys. Rev. E 65, 036104 (2002).

R. F. Hunter, K. de la Haye, J. M. Murray, J. Badham, T. W. Valente, M. Clarke, and F. Kee, “Social network interventions for health behaviours and outcomes: A systematic review and meta-analysis,” PLoS Med. 16, 1 (2019).

F. Morone and H. A. Makse, “Influence maximization in complex networks through optimal percolation,” Nature 524, 65 (2015).

I. Iacopini, G. Petri, A. Barrat, and V. Latora, “Simplicial models of social contagion,” Nat. Commun. 10, 1 (2019).

B. Jhun, M. Jo, and B. Kahng, “Simplicial seis model in scale-free uniform hypergraph,” J. Stat. Mech.: Theory Exp 2019, 123207 (2019).

G. F. de Arruda, G. Petri, and Y. Moreno, “Social contagion models on hypergraphs,” Phys. Rev. Res. 2, 023032 (2020).

L. V. McFarland and W. E. Stamm, “Review of clostridium difficile-associated diseases,” Am. J. Infect. Control 14, 99 (1986).

L. Hébert-Dufresne, A. Allard, J.-G. Young, and L. J. Dubé, “Global efficiency of local immunization on complex networks,” Sci. Rep. 3, 2171 (2013).

J. Grilli, G. Barabás, M. J. Michalska-Smith, and S. Allesina, “Higher-order interactions stabilize dynamics in competitive network models,” Nature 548, 210 (2017).

D. J. P. O’Sullivan, G. J. O’Keeffe, P. G. Fennell, and J. P. Gleeson, “Mathematical modeling of complex contagion on clustered networks,” Front. Phys. 3, 71 (2015).