Asymmetric percolation drives a double transition in sexual contact networks — Supporting Information —

Antoine Allard,[1] Benjamin M. Althouse,[2,3,4] Samuel V. Scarpino,[5] and Laurent Hébert-Dufresne[2,6,7]

[1] *Centre de Recerca Matemàtica, Edifici C, Campus Bellaterra, E-08193 Bellaterra (Barcelona), Spain*
[2] *Institute for Disease Modeling, Bellevue, WA, 98005, USA*
[3] *University of Washington, Seattle, WA, 98105, USA*
[4] *New Mexico State University, Las Cruces, NM, 88003, USA*
[5] *Department of Mathematics and Statistics and Complex Systems Center, University of Vermont, Burlington, VT, USA*
[6] *Santa Fe Institute, Santa Fe, NM, 87501, USA*
[7] *Department of Computer Science, University of Vermont, Burlington, VT, 05405, USA*

(Dated: June 27, 2017)

The analytical approach used in the main text is an adaptation of a more general formalism presented in Ref. [1]. We derive in more details its equations, validate its predictions with numerical simulations, and demonstrate some of the claims made in the main text. We also provide further details about the parameters used in the calculations. Finally, the general phenomenology and the results presented in the main text are further validated with simulations using realistic data.

The formalism presented in Ref. [1] is a generalization of the configuration model [2] in which nodes and connections are distinguished via types of nodes and stubs (i.e., half-links that are connected to create motifs). We consider a case in which there are 6 types of nodes based on the sex and sexual orientation of individuals (i.e., female/male and homo-/bi-/heterosexual). We denote the set of such types $\mathcal{N}$ and say that a fraction w_i of all nodes are of type i. Each node is assigned a number of contacts, k, independently of its type (i.e., all nodes have the same degree distribution $\{p_k\}_{k\geq 0}$), and links are created by randomly matching stubs respecting sexual orientations. Mathematically, this is encoded in the conditional probabilities $\{\alpha_{j|i}\}_{i,j\in\mathcal{N}}$ giving the probability that a link leaving a node of type i leads to a node of type j (see Sec. C for their numerical values). These probabilities are subjected to the constraints

$$\sum_{l\in\mathcal{N}} \alpha_{l|i} = 1 \ ; \qquad\qquad \alpha_{j|i} w_i = \alpha_{i|j} w_j \ , \tag{S1}$$

for every $i, j \in \mathcal{N}$. The latter constraint is due to the undirected nature of links and implies that there are as much links running from nodes of type i towards nodes of type j than in the opposite direction. Note that since nodes are only connected via links, the set of types of stubs is the same as the set of nodes and need not be considered explicitly (see Sec. IV of Ref. [1]).

Because links are created randomly, the probability that a node of type i and degree k has a set of $\boldsymbol{k} := \{k_j\}_{j\in\mathcal{N}}$ neighbors of each type is simply

$$P(\boldsymbol{k}|i,k) := \frac{k!}{\prod_{j\in\mathcal{N}} k_j!} \prod_{j\in\mathcal{N}} \left[\alpha_{j|i}\right]^{k_j} \tag{S2}$$

with the constraint that $\sum_{j\in\mathcal{N}} k_j = k$. Given that links running from nodes of type i to nodes of type j are independently removed with probability $1 - T_{ij}$, the probability that a node of type i and original degree k has $\boldsymbol{m}$ neighbors is

$$P(\boldsymbol{m}|i,k) := \sum_{\boldsymbol{k}\geq\boldsymbol{m}} P(\boldsymbol{k}|i,k) \prod_{j\in\mathcal{N}} \binom{k_j}{m_j} T_{ij}^{m_j} (1-T_{ij})^{k_j - m_j} \ , \tag{S3}$$

where the summation runs over $k_j \geq m_j$ for every $j \in \mathcal{N}$. The multivariate generating function associated with the joint degree distribution for nodes of type i is

$$\begin{aligned}
g_i(\boldsymbol{x}) &:= \sum_k p_k \sum_{\boldsymbol{m}} P(\boldsymbol{m}|i,k) \prod_{j\in\mathcal{N}} x_j^{m_j} \\
&= \sum_k p_k \sum_{\boldsymbol{m}} \sum_{\boldsymbol{k}\geq\boldsymbol{m}} P(\boldsymbol{k}|i,k) \prod_{j\in\mathcal{N}} \binom{k_j}{m_j} [T_{ij}x_j]^{m_j} (1-T_{ij})^{k_j-m_j} \\
&= \sum_k p_k \sum_{\boldsymbol{k}} P(\boldsymbol{k}|i,k) \prod_{j\in\mathcal{N}} \sum_{m_j=0}^{k_j} \binom{k_j}{m_j} [T_{ij}x_j]^{m_j} (1-T_{ij})^{k_j-m_j} \\
&= \sum_k p_k \sum_{\boldsymbol{k}} \frac{k!}{\prod_{j\in\mathcal{N}} k_j!} \prod_{j\in\mathcal{N}} \left[\alpha_{j|i}\right]^{k_j} \sum_{m_j=0}^{k_j} \binom{k_j}{m_j} [T_{ij}x_j]^{m_j} (1-T_{ij})^{k_j-m_j} \\
&= \sum_k p_k \sum_{\boldsymbol{k}} \frac{k!}{\prod_{j\in\mathcal{N}} k_j!} \prod_{j\in\mathcal{N}} \left[\alpha_{j|i}\right]^{k_j} \left[1-T_{ij}+T_{ij}x_j\right]^{k_j} \\
&= \sum_k p_k \left[\sum_{j\in\mathcal{N}} \alpha_{j|i} \left(1-T_{ij}+T_{ij}x_j\right)\right]^k
\end{aligned} \tag{S4}$$

by virtue of the multinomial theorem. The probability that a node of type i has $\boldsymbol{k}$ nodes *directly* accessible (i.e., via an existing link) is the coefficient of $g_i(\boldsymbol{x})$ in front of $\prod_{j\in\mathcal{N}} x_j^{k_j}$; hence we say that $g_i(\boldsymbol{x})$ *generates* the joint degree distribution for nodes of type i. It is worth mentioning that disregarding the types of nodes [i.e., setting $x_j = x$ and $T_{ij} = T$ for all $i, j \in \mathcal{N}$] in Eq. (S4) yields

$$g(x) = \sum_k p_k \left[1-T+Tx\right]^k \ , \tag{S5}$$

corresponding to the classical configuration model [2, 3]. In other words, even if distinguishing nodes into different types appears to force correlations in the way they are connected, our approach generates well-mixed networks that are structurally *indistinguishable* from networks generated with the configuration model and the same degree distribution.

To compute the size of the components requires to know the *excess* degree of nodes: the number of *new* nodes that can be reached from a node that has itself been reached via one of its links. The generating function associated with this distribution, $f_{li}(\boldsymbol{x})$, can readily be obtained from Eq. (S4) as

$$f_{li}(\boldsymbol{x}) := \left[\frac{\partial g_i(\boldsymbol{x})}{\partial x_l}\right]^{-1}_{\boldsymbol{x}=\boldsymbol{1}} \left[\frac{\partial g_i(\boldsymbol{x})}{\partial x_l}\right] = \sum_k \frac{k p_k}{\langle k\rangle} \left[\sum_{j\in\mathcal{N}} \alpha_{j|i} \left(1-T_{ij}+T_{ij}x_j\right)\right]^{k-1} . \tag{S6}$$

One striking result here is that $f_{li}(\boldsymbol{x})$ does not depend on the type of the node from which the node of type i has been reached. This is a direct consequence of the multinomial form of Eq. (S2). Consequently, we define

$$f_i(\boldsymbol{x}) := f_{li}(\boldsymbol{x}) \tag{S7}$$

for the remaining of the document.

A. Extensive (giant) component

With the generating function $g_i(\boldsymbol{x})$ and $f_i(\boldsymbol{x})$ in hand, we can already compute the size, composition and probability of the extensive, giant component in the limit of large size networks. Let us define u_i as the probability

that a neighbor of type i does not lead to the giant component (i.e., regardless of the type of the node whose neighbor is of type i). Because links are created by randomly matching stubs, we expect the networks generated by our model to be locally tree-like in the limit of large networks for a great variety of degree distributions. Consequently, the probabilities $\{u_i\}_{i\in\mathcal{N}}$ can be obtained via a simple self-consistent argument: if the neighbor of type i mentioned above does not lead to the extensive component, then neither should its other neighbors. Since the distribution of the number and of the type of these *other* neighbors is generated by $f_i(\boldsymbol{x})$, we directly have that the $\{u_i\}_{i\in\mathcal{N}}$ are the solutions of

$$u_i = f_i(\boldsymbol{u}) = \sum_k \frac{kp_k}{\langle k\rangle}\left[\sum_{j\in\mathcal{N}} \alpha_{j|i}\left(1 - T_{ij} + T_{ij}u_j\right)\right]^{k-1} \tag{S8}$$

for every $i \in \mathcal{N}$. Following the same line of thoughts, the probability that a randomly chosen node of type i leads to the giant component is simply the complement of the probability that neither of its neighbors leads to it, or in mathematical terms

$$P_i := 1 - g_i(\boldsymbol{u}) = 1 - \sum_k p_k\left[\sum_{j\in\mathcal{N}} \alpha_{j|i}\left(1 - T_{ij} + T_{ij}u_j\right)\right]^{k} . \tag{S9}$$

The probability that any given node leads to the giant component is therefore

$$P := \sum_{i\in\mathcal{N}} w_i P_i = \sum_{i\in\mathcal{N}} w_i[1 - g_i(\boldsymbol{u})] = 1 - \sum_{i\in\mathcal{N}} w_i \sum_k p_k\left[\sum_{j\in\mathcal{N}} \alpha_{j|i}\left(1 - T_{ij} + T_{ij}u_j\right)\right]^{k} . \tag{S10}$$

An important feature of asymmetric percolation on undirected networks – when links are undirected but the probability of existence of links is asymmetric (i.e., that $T_{ij} \neq T_{ji}$ from some i and j) – is that it is equivalent to percolation on semi-directed networks. This implies that in general, the relative size of the giant component, S, does not equal its probability of existence, P [1, 4]. In fact, while P is computed by looking at the fraction of nodes that *lead* to the extensive component (i.e., extensive in-component), its relative size is obtained by looking at the fraction of nodes that can be *reached* from the formers (i.e., extensive out-component). Since links in our model are originally undirected, the relative size of the extensive component is obtained by simply switching the probability of existence of links; T_{ij} becomes T_{ji} for all $i, j \in \mathcal{N}$. Indeed, as explained in Sec. D, a neighbor of type j can be reached from a node of type i with probability $T_{ij}T_{ji} + T_{ij}(1 - T_{ji}) = T_{ij}$, while the same neighbor of type j can reach the node of type i with probability $T_{ij}T_{ji} + T_{ji}(1 - T_{ij}) = T_{ji}$. In other words, switching the probability of existence of links effectively corresponds to switching from the out-degree to the in-degree. Let us define v_i as the probability that a node cannot be reached from the extensive component via a neighbor of type i, and using the same self-consistency argument as above, we find that it is the solution of

$$v_i = \sum_k \frac{kp_k}{\langle k\rangle}\left[\sum_{j\in\mathcal{N}} \alpha_{j|i}\left(1 - T_{ji} + T_{ji}v_j\right)\right]^{k-1} \tag{S11}$$

for every $i \in \mathcal{N}$. Similarly, the probability that a node of type i is not part of the extensive component, S_i, corresponds to the complement of the probability that none of its neighbors is part of it, or

$$S_i := 1 - \sum_k p_k\left[\sum_{j\in\mathcal{N}} \alpha_{j|i}\left(1 - T_{ji} + T_{ji}v_j\right)\right]^{k} , \tag{S12}$$

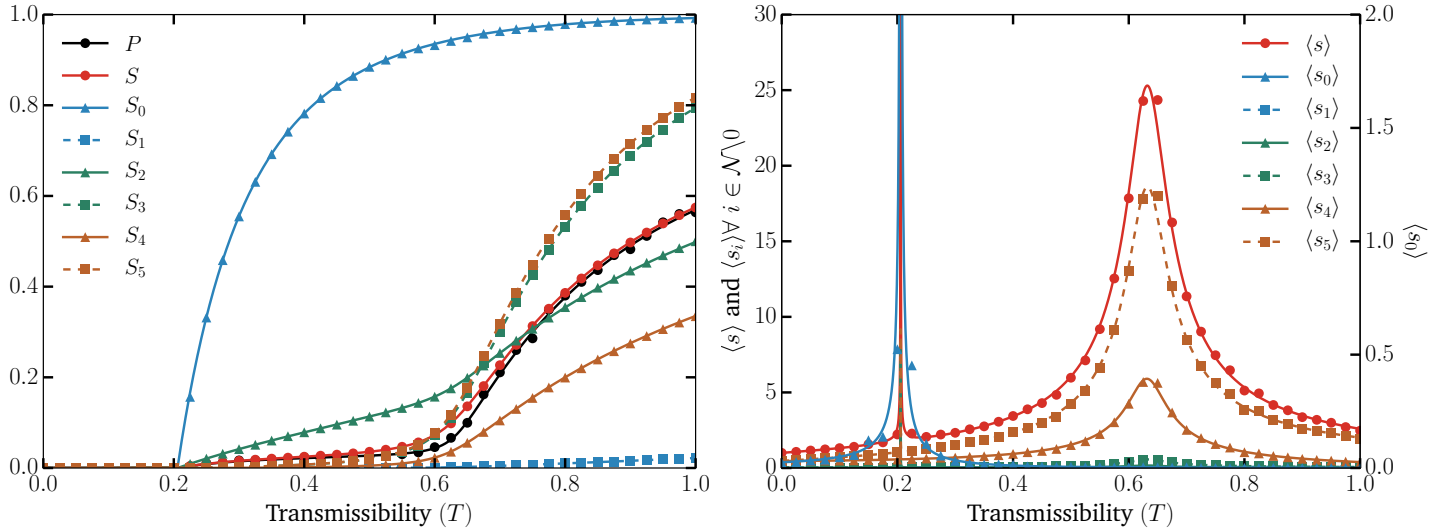

FIG. S1. **Validation of the analytical calculations.** The predictions (lines) of Eqs. (S8)–(S13) (left) and Eqs. (S21)–(S24) (right) are compared with the results obtained from numerical simulations (symbols). Note that $\langle s_0\rangle$ is shown using a different axis to facilitate the comparison between the theoretical prediction and the results of the numerical simulations. The asymmetry parameter is $a = 10$; details on the other parameters and numerical simualtions are given in Secs. C and D.

and the relative size of the extensive component is simply the fraction of nodes that can be reached from other nodes in it, or

$$S := \sum_{i\in\mathcal{N}} w_i S_i = 1 - \sum_{i\in\mathcal{N}} w_i \sum_k p_k \left[\sum_{j\in\mathcal{N}} \alpha_{j|i}\left(1 - T_{ji} + T_{ji}v_j\right)\right]^k . \tag{S13}$$

Since the right-hand side of Eqs. (S8) and (S11) are convex and normalized polynomials, the values of $\{u_i\}_{i\in\mathcal{N}}$ and $\{v_i\}_{i\in\mathcal{N}}$ can be obtained by iterating Eqs. (S8) and (S11) from an initial condition in $[0,1)^{|\mathcal{N}|}$ until convergence. The predictions of Eqs. (S8)–(S13) are validated on Fig. S1.

B. Non-extensive (small) components

The distribution of the composition of the small, non-extensive components can be computed in a similar fashion. Let us define the probability generating function (pgf) $H_i(\boldsymbol{x})$ whose coefficients correspond to the probability that a neighbor of type i leads to a small component of a given composition (i.e., the number of nodes of type j is given by the exponent of x_j). Invoking the same self-consistency argument as above, the pgfs are the solution of

$$H_i(\boldsymbol{x}) = x_i \sum_k \frac{kp_k}{\langle k\rangle}\left[\sum_{j\in\mathcal{N}} \alpha_{j|i}[1 - T_{ij} + T_{ij}H_j(\boldsymbol{x})]\right]^{k-1} . \tag{S14}$$

where the extra x_i has been added to account for the neighbor of type i itself. Similarly, the small component that can be reached from a node of type i is given by

$$K_i(\boldsymbol{x}) := x_i \sum_k p_k \left[\sum_{j\in\mathcal{N}} \alpha_{j|i}[1 - T_{ij} + T_{ij}H_j(\boldsymbol{x})]\right]^k . \tag{S15}$$

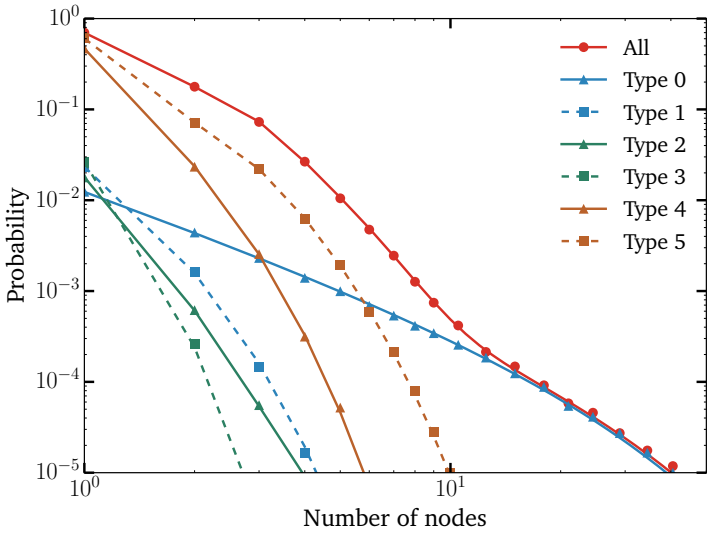

FIG. S2. **Validation of the analytical calculations.** The marginal distributions for the size of the small components predicted by Eqs. (S18)–(S20) (lines) are compared with the results obtained from numerical simulations (symbols). The asymmetry parameter is $a = 10$; details on the other parameters and numerical simualtions are given in Secs. C and D.

The distribution of the composition of the small components is

$$K(\boldsymbol{x}) := \sum_{i\in\mathcal{N}} w_i K_i(\boldsymbol{x}) = \sum_{i\in\mathcal{N}} w_i x_i \sum_k p_k \left[\sum_{j\in\mathcal{N}} \alpha_{j|i}[1 - T_{ij} + T_{ij} H_j(\boldsymbol{x})]\right]^k . \tag{S16}$$

It is worth noting that whenever $S > 0$, the distribution generated by $K(\boldsymbol{x})$ is no longer normalized, $K(\mathbf{1}) < 1$, since there is a non-zero probability that a given node does not lead to a finite component. In fact, a comparison of Eqs. (S8)–(S10) with (S14)–(S16) yields the following relations

$$u_i = H_i(\mathbf{1}) \; ; \qquad P_i = 1 - K_i(\mathbf{1}) \; ; \qquad P = 1 - K(\mathbf{1}) \; . \tag{S17}$$

for every $i \in \mathcal{N}$.

Although the pgf $K(\boldsymbol{x})$ generates the joint distribution of the number of nodes of each type in small components, we are typically interested in its marginal distributions (i.e., the distribution of the number of nodes of a given type). To obtain the distribution of the number of nodes of a given type, say type i, we simply set $x_j = \delta_{ij} z$ for all $j \in \mathcal{N}$ in Eqs. (S14)–(S16). Similarly, if we are interested in the size of the components, that is the number of nodes regardless of their type, we simply set $x_j = z$ for all $j \in \mathcal{N}$. The marginal distribution for the number of nodes in a component, $P(s)$, is then

$$P(s) = \left[\frac{1}{s!}\frac{\partial^s}{\partial z^s}\frac{K(z)}{K(1)}\right]_{z=0} . \tag{S18}$$

From Cauchy's integral formula, this last expression can be written in terms of a path integral in the complex plane

$$P(s) = \frac{1}{2i\pi K(1)} \oint_\Gamma \frac{K(z)}{z^{s+1}} dz \; , \tag{S19}$$

where Γ is a closed circular path of radius r centered at the origin and in which $K(z)$ is analytic. Substituting $z = re^{2i\pi\phi}$ and discretizing the integral in $s_{\max}$ equally spaced points, this last expression becomes

$$\begin{aligned} P(s) &= \frac{1}{2i\pi K(1)} \oint_{\Gamma} \frac{K(z)}{z^{s+1}} dz \\ &= \frac{1}{r^s K(1)} \int_0^1 K\big(re^{2i\pi\phi}\big) e^{-2i\pi s\phi} d\phi \\ &\simeq \frac{1}{s_{\max} r^s K(1)} \sum_{n=0}^{s_{\max}-1} K\big(re^{2i\pi n/s_{\max}}\big) e^{-2i\pi sn/s_{\max}} \ , \end{aligned} \tag{S20}$$

where we identify the inverse discrete Fourier transform of the sequence $\{K\big(re^{2i\pi n/s_{\max}}\big)\}$ with $n = 0, \ldots, s_{\max} - 1$. Although the exact form of $K(z)$ is unknown, it is nevertheless possible to evaluate the sequence $\{K\big(re^{2i\pi n/s_{\max}}\big)\}$ by solving Eqs. (S14)–(S16) and then to quickly obtain $P(s)$ for $s < s_{\max}$ with a single pass of the FFT algorithm. This method is validated on Fig. S2.

The emergence of a critical cluster at the phase transition is associated with the divergence of the average size of the small components

$$\langle s \rangle = \sum_{i \in \mathcal{N}} \langle s_i \rangle \ , \tag{S21}$$

which can be computed from the marginal distributions obtained with the method explained above. However, since $P(s) \sim s^{-3/2}$ at the phase transition [2], calculating $\langle s_i \rangle$ near the percolation threshold requires large values of $s_{\max}$ which make this method inefficient in practice. Luckily, there is another way to compute the average number of nodes in small components without having to compute the marginal distribution beforehand. Since the moments of a distribution can be obtained via a suitable differentiation of its associated pgf, the average number of nodes of type i in small components can also be written as

$$\langle s_i \rangle = \frac{1}{K(\mathbf{1})} \left. \frac{dK(\boldsymbol{x})}{dx_i} \right|_{\boldsymbol{x}=\mathbf{1}} \ . \tag{S22}$$

Substituting Eq. (S16) in this equation yields

$$\langle s_i \rangle = \frac{w_i g_i(\boldsymbol{u})}{K(1)} + \frac{\langle k \rangle}{K(1)} \sum_{j,l \in \mathcal{N}} w_j u_j \alpha_{l|j} T_{jl} \frac{\partial H_l(\mathbf{1})}{\partial x_i} \ , \tag{S23}$$

where the equation for $\frac{\partial H_l(\mathbf{1})}{\partial x_i}$ is obtained via Eq. (S14)

$$\frac{\partial H_l(\mathbf{1})}{\partial x_i} = \delta_{il} f_l(\boldsymbol{u}) + \left[\sum_k \frac{k(k-1)p_k}{\langle k \rangle} \left(\sum_{j \in \mathcal{N}} \alpha_{j|l} \left[1 - T_{lj} + T_{lj} u_j \right] \right)^{k-2} \right] \times \left[\sum_{j \in \mathcal{N}} \alpha_{j|l} T_{lj} \frac{\partial H_j(\mathbf{1})}{\partial x_i} \right] \ , \tag{S24}$$

for every $i, l \in \mathcal{N}$. Consequently, the average number of nodes of each type in small components can be obtained by solving Eq. (S24) for every combination of $i, l \in \mathcal{N}$ and then injecting the solutions in Eq. (S23). This approach is validated on Fig. S1.

C. Choice of parameters

The proportion of each type of nodes, $\{w_i\}_{i \in \mathcal{N}}$ (see Table S1), used for the calculations are conservative yet realistic estimates extrapolated from the data provided in Refs. [5–7] with respect to the conclusions discussed in

Description	Homo. males.	Homo. females	Bi. males	Bi. females	Hetero. males	Hetero. females
i	0	1	2	3	4	5
w_i	0.025	0.025	0.015	0.015	0.46	0.46

TABLE S1. Description of the node types and their relative proportion used in all calculations and numerical simulations.

the main text. The values chosen for the asymmetry parameter a (i.e., 3 and 10), are also either quite conservative (3) or in line (10) with the naive estimates made from the infectious periods of males and females (see Introduction in the main text). Moreover, although there is evidence that the probability of HIV transmission differs by sexual act and insertive/receptor status (i.e., receptive anal sex could be as much as 20 times more likely to transmit HIV than insertive vaginal sex, see Ref. [8]), we are not aware of any strong evidence suggesting that the probability of transmission of ZIKV is different based on whether the *receptor* was male or female. We therefore opted for a conservative approach and chose to use the same probability of transmission regardless of the gender of the receiving partner (i.e., T_{ij} does not depend on j), that is

$$T_{ij} = \begin{cases} T & \text{if } i \text{ is male} \\ T/a & \text{if } i \text{ is female} \end{cases} \tag{S25}$$

for every $j \in \mathcal{N}$. The use of fixed probabilities of transmission relies implicitly in the so-called i.i.d. assumption introduced in Ref. [3] which ignores, due to coarse-graining, some correlations and heterogeneity associated with the distribution of infectious periods. To understand the implications of the independent identically distributed (i.i.d.) assumption, let us consider an individual of type i that is infectious for a period τ and that interacts with one of its neighbors of type j at a rate β. The probability that this individual *never* infects its neighbor is

$$\lim_{\delta t \to 0} (1 - \beta \delta t)^{\tau/\delta t} = e^{-\beta \tau} \ . \tag{S26}$$

Assuming that τ and β are i.i.d. random variables respectively drawn from the distributions $P_i(\tau)$ and $Q_{ij}(\beta)$, the probability that *any* individual of type i *eventually* infects one of its neighbors of type j is

$$T_{ij} = 1 - \iint e^{-\beta \tau} P_i(\tau) Q_{ij}(\beta) d\tau d\beta \ . \tag{S27}$$

The i.i.d. assumption therefore implies that each infection event is independent for the other infection events, even if these infection events all come from the same individual. However, although correlations in the infectious period at the individual level are not taken into account, they are taken into account at the gender level since the probability of transmission, T_{ij}, corresponds to the probability that a node of type i will eventually infect its neighbor of type j and, as such, is obtained by averaging over a distribution of infectious period, $P_i(\tau)$, that depends explicitly on i. Since the asymmetry in the probability of transmission between men and women is mainly due to men having longer infectious periods than women, the use of node types is expected to capture the relevant heterogeneity in the infectious period with regards to the double transition phenomenology. In other words, the i.i.d. assumption coupled with multiple types of nodes (i.e., the use of T_{ij} instead of only one unique probability of transmission T) is sufficient to take into account that men are more likely to transmit ZIKV than women due to their longer infectious periods. Adapting the frameworks developed in Refs. [9, 10] to multiple types of nodes would get rid of the i.i.d. asumption but is not expected to alter the qualitative aspect of the double transition phenomenology.

The values of $\{\alpha_{j|i}\}_{i,j \in \mathcal{N}}$ are chosen to respect as much as possible the proportion of each type of nodes, $\{w_i\}_{i \in \mathcal{N}}$ (see Table S1), given the constraints of Eqs. (S1). Taking the nodes of type 2 as a starting point, we set

$$\alpha_{0|2} = \frac{w_0}{w_0 + w_2 + w_3 + w_5} \ ; \qquad \alpha_{2|2} = \frac{w_2}{w_0 + w_2 + w_3 + w_5}$$
$$\alpha_{3|2} = \frac{w_3}{w_0 + w_2 + w_3 + w_5} \ ; \qquad \alpha_{5|2} = \frac{w_5}{w_0 + w_2 + w_3 + w_5} \ ,$$

$i \setminus j$	0	1	2	3	4	5
0	0.971	0.000	0.029	0.000	0.000	0.000
1	0.000	0.971	0.000	0.029	0.000	0.000
2	0.049	0.000	0.029	0.029	0.000	0.893
3	0.000	0.049	0.029	0.029	0.893	0.000
4	0.000	0.000	0.000	0.029	0.000	0.971
5	0.000	0.000	0.029	0.000	0.971	0.000

TABLE S2. Values of $\alpha_{j|i}$ used in all calculations and numerical simulations.

from which we can readily obtain the following from Eqs. (S1)

$$\alpha_{2|0} = \frac{w_2\alpha_{0|2}}{w_0} \; ; \qquad \alpha_{2|3} = \frac{w_2\alpha_{3|2}}{w_3} \; ; \qquad \alpha_{2|5} = \frac{w_2\alpha_{5|2}}{w_5} \; : \qquad \alpha_{0|0} = 1 - \alpha_{2|0}$$
$$\alpha_{4|5} = 1 - \alpha_{2|5} \; ; \qquad \alpha_{5|4} = \frac{w_5\alpha_{4|5}}{w_4} \; ; \qquad \alpha_{3|4} = 1 - \alpha_{5|4} \; ; \qquad \alpha_{4|3} = \frac{w_4\alpha_{3|4}}{w_3} \; .$$

Taking a similar point of view for nodes of type 3, but taking into account that $\alpha_{2|3}$ and $\alpha_{4|3}$ are already known, we set

$$\alpha_{1|3} = \frac{[1 - (\alpha_{2|3} + \alpha_{4|3})]w_1}{w_1 + w_3} \; ; \qquad \alpha_{3|3} = \frac{[1 - (\alpha_{2|3} + \alpha_{4|3})]w_3}{w_1 + w_3} \; ,$$

from which we finally obtain

$$\alpha_{3|1} = \frac{w_3\alpha_{1|3}}{w_1} \; ; \qquad \alpha_{1|1} = 1 - \alpha_{3|1} \; .$$

All other probabilities are set to zero. The values obtained using the parameters given in Table S1 are given in Table S2.

D. Details on the numerical simulations

The analytical predictions of the model are validated by comparing them with results obtained from numerical simulations in which networks are generated according to the stub matching scheme of Ref. [1]. More precisely, 10^6 nodes are first assigned a number of stubs, corresponding to their degree, drawn from a Poisson distribution

$$p_k = \frac{e^{-\langle k \rangle}\langle k \rangle^k}{k!} \tag{S28}$$

where $k \geq 0$ and with an average degree $\langle k \rangle = 5$. Stubs are then paired up to form links as follows.

1. The stubs assigned to each node are gathered in six lists of unmatched stubs (one list for each node type).
2. With probability

$$R_{ij} = \frac{w_i\alpha_{j|i}}{1 + \delta_{ij}} \left[\sum_{m \geq n \in \mathcal{N}} \frac{w_m\alpha_{n|m}}{1 + \delta_{mn}} \right]^{-1} \; , \tag{S29}$$

one stub attached to a node of type i and one stub attached to a node of type j are randomly chosen and are matched to form a link. These stubs are then removed from their respective list.

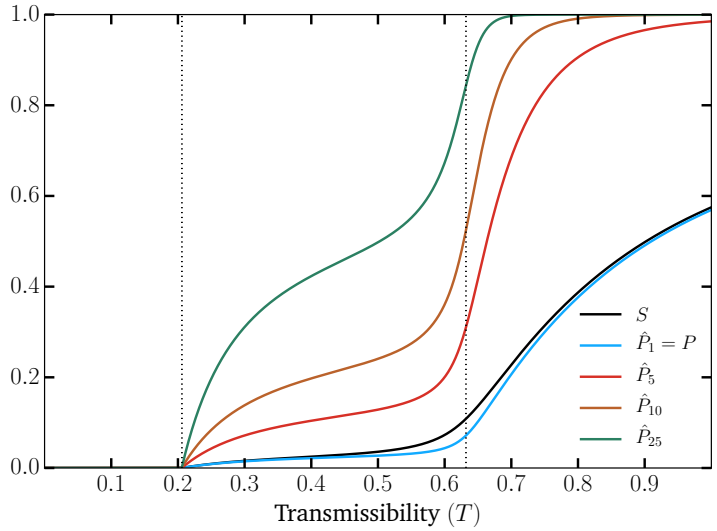

FIG. S3. **Effect of multiple seeds.** Behavior of Eq. (S30) for various values of N using the same parameters as in Fig. S1. The two vertical dotted lines show the thresholds for the the two transitions as introduced in the main text.

3. Step 2 is repeated until every list is *effectively* empty (i.e., some stubs may remained unmatched due to statistical fluctuations but their number is typically small enough such that the observables measured on the generated networks are not affected).

The generated *undirected* network is then transformed into a *semi-directed* network as follows. A link between a node of type i and a node of type j

- remains undirected with probability $T_{ij}T_{ji}$;
- becomes directed from the node of type i to the node of type j ($i \to j$) with probability $T_{ij}(1 - T_{ji})$;
- becomes directed from the node of type j to the node of type i ($j \to i$) with probability $T_{ji}(1 - T_{ij})$;
- is removed with probability $(1 - T_{ij})(1 - T_{ji})$.

This procedure is repeated for every link. The probability that a node of type i starts an epidemic, P_i, then corresponds to the fraction of nodes of type i that are in the extensive in-component, and the probability that a node of type i is infected during an epidemic, S_i, corresponds to the fraction of nodes of type i that are in the extensive out-component. Similarly, the average composition of outbreaks, $\langle s_i \rangle$ $\forall i \in \mathcal{N}$, are obtained by looking at the composition of non-extensive out-components.

Since the observables measured on Figs. S1 and S2 are self-averaging properties, the number of semi-directed networks over which they were averaged was chosen to be large enough so that the stochastic fluctuations were reduced to a satisfactory level (about 5000 for each value of T).

E. Effect of multiple seeds

The effect of having multiple seeds due to different individuals being infected by mosquitoes can be investigated using our formalism. In Sec. A, we define P as the probability that a random individual sparks an epidemic of size S. If we assume that N such individuals have been independently infected by mosquitoes (i.e., they are sufficiently far apart in the contact network), the probability that at least one of them sparks an epidemic is

$$\hat{P}_N = 1 - (1 - P)^N \quad . \tag{S30}$$

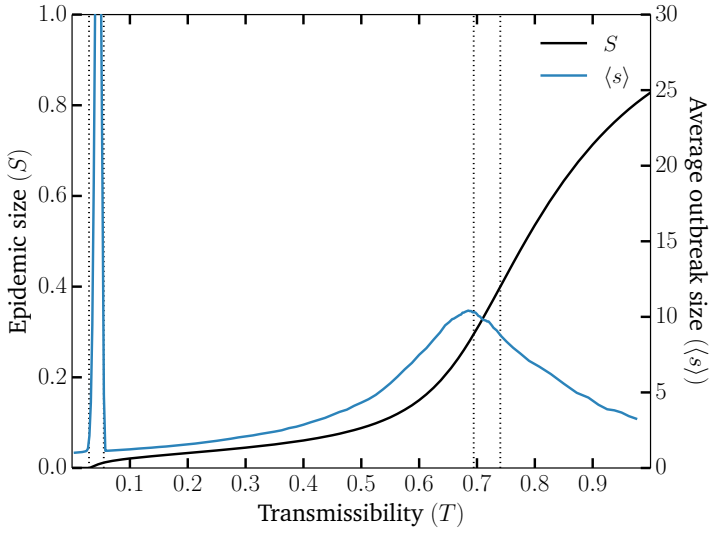

FIG. S4. **Simulations using realistic data.** The size of the extensive component, S, and the average size of microscopic outbreaks, $\langle s \rangle$ are shown for the random network ensemble defined in Sec. F. These results confirm the general phenomenology presented in the main text as well as the qualitative accuracy of the naive estimates of the upper and lower bounds of the thresholds (black dotted vertical lines). Note that the distribution of the size of the small components follows a power law on many orders of magnitude at the second transition. This obfuscates the position of the peak in practice for finite-size networks. Based on Fig. 2 in the main text, we expect the peak to be slightly at the right of the line at $T \simeq 0.69$.

Moreover, since a fraction $1-P$ of all seeds are expected to *not* spark a macroscopic epidemic, N seeds are expected to infect $(1-P)N\langle s \rangle$ individuals in microscopic outbreaks. As N increases, we therefore expect a higher probability of observing a macroscopic epidemic as well as more individuals infected in microscopic outbreaks. Figure S3 shows the behavior of the quantity $\hat{P}_N$ for the parameters used in Fig. S1. We see that although multiple seeds tend to smooth the transition between the two thresholds with respect to the probability of sparking an epidemic, a similar phenomenology is observed. More importantly, as shown in Ref. [11], the size of the epidemic is unaffected by a finite number of seeds in the limit of large network size. In other words, having multiple seeds increases the likelihood of reaching the extensive out-component, but it does not change its size. While this only maps to vector transmission once a steady-state is achieved, it does suggest that the general phenomenology presented in the main text is valid with or without vector transmission.

F. Simulations using realistic data

To further support our results, we simulated ZIKV transmission dynamics on a random network of sexual contacts where the distribution of contacts are drawn from Ref. [5] (for homosexual contacts) and Ref. [7] (for heterosexual contacts) but only using individuals with more than one sexual partner (as a rough approximation of the sexually active population). A population of 10^6 individuals is assumed to be equally parted between men and women, of which 5% are homosexuals, 3% are bisexuals and 92% are heterosexuals. We assume very few individuals are infected by the mosquito vector, but they can cause outbreaks through sexual transmission where men transmit with probability T and women with probability T/a. Using $a = 2$, Fig. S4 shows that increasing heterogeneity in the distribution of the degrees does not alter the phenomenology presented in the main text. Moreover, we see that the "naive" estimates described in the caption of Fig. 2 in the main text remain good

indicators for the location of the two epidemic thresholds.

[1] A. Allard, L. Hébert-Dufresne, J.-G. Young, and L. J. Dubé, "General and exact approach to percolation on random graphs," Phys. Rev. E **92**, 062807 (2015).

[2] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," Phys. Rev. E **64**, 026118 (2001).

[3] M. E. J. Newman, "Spread of epidemic disease on networks," Phys. Rev. E **66**, 016128 (2002).

[4] A. Allard, P.-A. Noël, L. J. Dubé, and B. Pourbohloul, "Heterogeneous bond percolation on multitype networks with an application to epidemic dynamics," Phys. Rev. E **79**, 036113 (2009).

[5] A. E. Grulich, R. O. Visser, A. M. A. Smith, C. E. Rissel, and J. Richters, "Sex in Australia: Homosexual experience and recent homosexual encounters," Aust. N. Z. J. Public Health **27**, 155–163 (2003).

[6] A. K. Baumle, "Introduction: The Demography of Sexuality," in *International Handbook on the Demography of Sexuality*, International Handbooks of Population, Vol. 5, edited by A. K. Baumle (Springer, 2013) pp. 3–9.

[7] A. Chandra, C. E. Copen, and W. D. Mosher, "Sexual Behavior, Sexual Attraction, and Sexual Identity in the United States: Data from the 2006-2010 National Survey of Family Growth," in *International Handbook on the Demography of Sexuality*, International Handbooks of Population, Vol. 5, edited by A. K. Baumle (Springer, 2013) pp. 45–66.

[8] M.-C. Boily, R. F. Baggaley, L. Wang, B. Masse, R. G. White, R. J. Hayes, and M. Alary, "Heterosexual risk of HIV-1 infection per sexual act: Systematic review and meta-analysis of observational studies," Lancet Infect. Dis. **9**, 118–129 (2009).

[9] E. Kenah and J. M. Robins, "Second look at the spread of epidemics on networks," Phys. Rev. E **76**, 036113 (2007).

[10] J. C. Miller, "Epidemic size and probability in populations with heterogeneous infectivity and susceptibility," Phys. Rev. E **76**, 010101 (2007).

[11] L. A. Meyers, B. Pourbohloul, M. E. J. Newman, D. M. Skowronski, and R. C. Brunham, "Network theory and SARS: Predicting outbreak diversity." J. Theor. Biol. **232**, 71–81 (2005).