Direct, physically motivated derivation of triggering probabilities for spreading processes on generalized random networks

Peter Sheridan Dodds,[1,2,*] Kameron Decker Harris,[1,2,†] and Joshua L. Payne[3,‡]
[1] *Department of Mathematics & Statistics, The University of Vermont, Burlington, VT 05401.*
[2] *Complex Systems Center & the Vermont Advanced Computing Center, The University of Vermont, Burlington, VT 05401.*
[3] *Computational Genetics Laboratory, Dartmouth College, Hanover, NH 03755*

(Dated: August 30, 2011)

We derive a general expression for the probability of global spreading starting from a single infected seed for contagion processes acting on generalized, correlated random networks. We employ a simple probabilistic argument that encodes the spreading mechanism in an intuitive, physical fashion. We use our approach to directly and systematically obtain triggering probabilities for contagion processes acting on a series of interrelated random network families.

PACS numbers: 89.75.Hc,64.60.aq,64.60.Bd,87.23.Ge

I. INTRODUCTION

Spreading is a pervasive dynamic phenomenon, ranging in form from simple physical diffusion to the complexities of socio-cultural dispersion and interaction of ideas and beliefs [1–11]. Successful spreading in systems may manifest as an expanding front, such as in the spread of disease through medieval Europe [12], or through inherent or revealed networks, such as in pandemics in the modern era of global travel [13]. Here, we focus on spreading processes operating on generalized random networks, which have proven over the last decade to be illustrative of spreading on real networks and at the same time to be analytically tractable [3, 14–24].

In contributing to the wealth of already known results for random network contagion model, we make two main advances here. First, we obtain, in the most general terms possible, an expression for the probability of global spreading from a single seed for a broad range of contagion processes acting on generalized, correlated random networks. By global spreading we mean a non-zero fraction of nodes in an infinite network are eventually infected. Second, we use an argument that is physically motivated and direct. Existing approaches rely on a range of mathematical techniques, such as probability generating functions [18, 25, 26], which, while being entirely successful in determining spreading probabilities as well as many other quantities, involve an obfuscation of the underlying physical mechanisms.

The present paper is a companion to our earlier work where we derived a general condition for the possibility (rather than probability) of global spreading for single-seed contagion processes acting on random networks [27].

We structure our paper as follows. In Sec. II, we define the broadest class of correlated random networks allowing for directed and undirected edges and arbitrary node and edge properties. In Sec. III, we define the general class of contagion processes that our treatment can encompass. In Sec. IV, we compute the probability that seeding a node of a given type generates a global spreading event. We use our formalism for six interrelated random network families with general contagion processes acting on them in Sec. V, and we offer some concluding remarks in Sec. VI.

II. GENERALIZED RANDOM NETWORKS

Our theoretical treatment builds on a formalism we introduce here for representing generalized random networks, an expansion of what we used in our connected, earlier work [27]. We depict the essential features of a random network with the possibility of directed edges in Fig. 1. The most basic elements of networks are nodes and edges, and here we allow the following features encoded in two parameters:

- ν represents a set of arbitrary node characteristics such as node degree, node age, susceptibility to a given disease or message, etc.; and
- λ represents a set of arbitrary edge characteristics such as direction, age, strength, conductance, etc. Since edges may have directional aspects, we must always be clear that edges are defined within the context of the nodes they connect. The variable λ is defined with respect to a given direction of travel along a specific edge. We hence also use the notation $\bar{\lambda}$ to indicate the edge's measured characteristics when travelled in the opposite direction. In other words, if in moving in one direction along an edge between two nodes labelled, say, i_1 and i_2 respectively, we observe the edge is of type λ, then in travelling in the reverse direction from node i_2 to i_1, the edge will be of type $\bar{\lambda}$. A simple example is unweighted, directed networks where we can travel with or against the direction of each edge.

*Electronic address: peter.dodds@uvm.edu
†Electronic address: kameron.harris@uvm.edu
‡Electronic address: joshua.payne@dartmouth.edu

Typeset by REVTEX

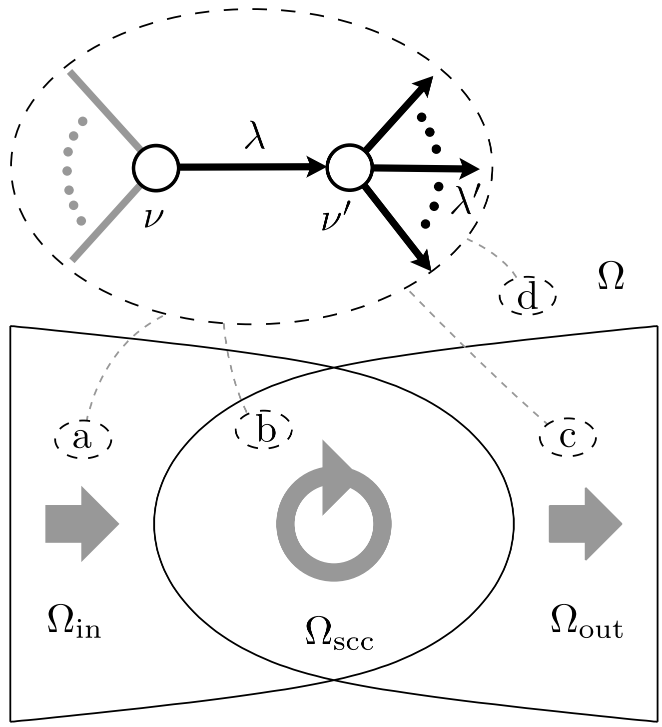

FIG. 1: Schematic showing the configuration of the potential triggering node subnetwork using the present work's formalism for generalized random networks described in Sec. II, and the basic form of a random network with directed edges and a giant component. The ellipses labelled a–d show four possible locations of the subnetwork in the overall network Ω. Global spreading events can be successfully generated only if the subnetwork is part of the giant in-component $\Omega_{\rm in}$, either within or outside of the giant strongly connected component $\Omega_{\rm scc}$ (ellipses a and b). No spreading is possible if the subnetwork is instead part of the giant out-component outside of the strongly connected component ($\Omega_{\rm out}/\Omega_{\rm scc}$, ellipse c) or outside of all three giant components (ellipse d).

We take all variables to be discrete, a choice that does not limit us when considering applications for real world networks. We denote the sets of all node and edge types by $\mathcal{N}$ and $\mathcal{L}$, the entire network by Ω, and the set of edges incident to a node of type ν by Λ_ν.

To define a random network with arbitrary node-edge-node correlations, we need to specify a number of interrelated probabilities, and these must further satisfy certain restrictions and detailed balance equations [14]. First and overall, we have the node and edge distributions $\mathbf{Pr}(\nu)$ and $\mathbf{Pr}(\lambda)$, where in randomly selecting an edge we must also randomly choose which direction to traverse it. Note that we immediately must have the evident restriction $\mathbf{Pr}(\lambda) = \mathbf{Pr}(\bar{\lambda})$, and of course the basic normalizations $\sum_{\nu\in\mathcal{N}}\mathbf{Pr}(\nu) = 1$ and $\sum_{\lambda\in\mathcal{L}}\mathbf{Pr}(\lambda) = 1$.

Next, we need $\mathbf{Pr}(\nu\lambda)$ which we define as the probability that in randomly choosing an edge and traversing it in a randomly chosen direction, we find it is of type λ and that we are travelling away from a node of type ν.

Finally, we encode correlations via the transition probability $\mathbf{Pr}(\nu'|\nu\lambda)$ which is the probability that we reach a type ν' node, given that we are following a type λ edge away from a type ν node.

We are now forced to connect and constrain the probabilities $\mathbf{Pr}(\nu\lambda)$ and $\mathbf{Pr}(\nu'|\nu\lambda)$ as follows. Consider $\mathbf{Pr}(\nu\lambda\nu')$ which we define as the probability a randomly selected edge is of type λ and runs between a type ν node and type ν' node (corresponding to the subnetwork in Fig. 1). This quantity depends on our two probabilities as $\mathbf{Pr}(\nu\lambda\nu') = \mathbf{Pr}(\nu'|\nu\lambda)\mathbf{Pr}(\nu\lambda)$. Now, because we may traverse edges in either direction with equal likelihood, we must also have $\mathbf{Pr}(\nu\lambda\nu') = \mathbf{Pr}(\nu'\bar{\lambda}\nu)$. We therefore arrive at the detailed balance condition:

$$\underbrace{\mathbf{Pr}(\nu'|\nu\lambda)\mathbf{Pr}(\nu\lambda)}_{\mathbf{Pr}(\nu\lambda\nu')} = \underbrace{\mathbf{Pr}(\nu|\nu'\bar{\lambda})\mathbf{Pr}(\nu'\bar{\lambda})}_{\mathbf{Pr}(\nu'\bar{\lambda}\nu)}. \qquad (1)$$

Before considering contagion processes, we recall the well-known typical macroscopic 'bow-tie' form of random networks with directed edges [14, 18, 28], given that a giant component is present. As shown in Fig. 1, there are three giant components of functional importance: (1) the giant strongly connected component, $\Omega_{\rm scc}$, within which any pair of nodes can be connected via a path of directed and/or undirected edges, traversing the directed ones; (2) the giant in-component $\Omega_{\rm in}$, the set of all nodes from which paths lead to $\Omega_{\rm scc}$ (n.b., $\Omega_{\rm scc} \subset \Omega_{\rm in}$); and (3) the giant out-component $\Omega_{\rm out}$, the set of all nodes which can be reached along directed paths starting from a node in $\Omega_{\rm in}$ (n.b., $\Omega_{\rm scc} \subset \Omega_{\rm out}$). By definition, we have that $\Omega_{\rm scc} = \Omega_{\rm in} \cap \Omega_{\rm out}$. Any global spreading event must begin from a seed in the giant in-component, and can at most spread to the giant out-component $\Omega_{\rm out}$.

III. GENERALIZED CONTAGION PROCESS

We consider contagion processes where the probability of a node's infection may depend in any fashion on the current states of its neighbors, potentially resembling phenomena ranging from the spread of infectious diseases to socially-transmitted behaviors [19, 29–31]. Since we are interested in the probability of spreading, we can capitalize on the fact that random networks are locally pure branching structures. We therefore need to know only what the probability of infection is for a type ν' node given a single neighbor of type ν is infected, whose influence is felt along a type λ edge. We write this probability as $B_{\nu\lambda\nu'}$. Time is removed from this quantity, as we need to know only the probability of eventual infection. Disease spreading models with recovery [27, 31] are included, as are threshold models inspired by social contagion [19, 30].

IV. TRIGGERING PROBABILITIES

We define $Q_{\nu\lambda}$ to be the probability that seeding a type ν node generates a global spreading event along an

Network:	Edge Triggering Probability:	Node Triggering Probability, Q:
I. Undirected, Uncorrelated	$Q_{**} = \sum_{k'_{\rm u}} P^{(\rm u)}(k'_{\rm u}\,\vert\,*) B_{**k'_{\rm u}} \left[1-(1-Q_{**})^{k'_{\rm u}-1}\right]$	$\sum_{k'_{\rm u}} \mathbf{Pr}(k'_{\rm u}) \left[1-(1-Q_{**})^{k'_{\rm u}}\right]$
II. Directed, Uncorrelated	$Q_{**} = \sum_{k'_{\rm i},k'_{\rm o}} P^{(\rm u)}(k'_{\rm i},k'_{\rm o}\vert\,*) B_{**k'_{\rm i}} \left[1-(1-Q_{**})^{k'_{\rm o}}\right]$	$\sum_{k'_{\rm i},k'_{\rm o}} \mathbf{Pr}(k'_{\rm i},k'_{\rm o}) \left[1-(1-Q_{**})^{k'_{\rm o}}\right]$
III. Mixed Directed and Undirected, Uncorrelated	$Q_{*u} = \sum_{\vec{k}'} P^{(\rm u)}(\vec{k}'\vert\,*) B_{**\vec{k}'} \left[1-(1-Q_{*u})^{k'_{\rm u}-1}(1-Q_{*o})^{k'_{\rm o}}\right]$ $Q_{*o} = \sum_{\vec{k}'} P^{(\rm i)}(\vec{k}'\vert\,*) B_{**\vec{k}'} \left[1-(1-Q_{*u})^{k'_{\rm u}}(1-Q_{*o})^{k'_{\rm o}}\right]$	$\sum_{\vec{k}'} \mathbf{Pr}(\vec{k}') \left[1-(1-Q_{*u})^{k'_{\rm u}}(1-Q_{*o})^{k'_{\rm o}}\right]$
IV. Undirected, Correlated	$Q_{k_{\rm u}*} = \sum_{k'_{\rm u}} P^{(\rm u)}(k'_{\rm u}\,\vert\,k_{\rm u}) B_{**k'_{\rm u}} \left[1-(1-Q_{k'_{\rm u}*})^{k'_{\rm u}-1}\right]$	$\sum_{k'_{\rm u}} \mathbf{Pr}(k'_{\rm u}) \left[1-(1-Q_{k'_{\rm u}*})^{k'_{\rm u}}\right]$
V. Directed, Correlated	$Q_{k_{\rm i}k_{\rm o},*} = \sum_{k'_{\rm i},k'_{\rm o}} P^{(\rm u)}(k'_{\rm i},k'_{\rm o}\vert\,k_{\rm i},k_{\rm o}) B_{**k'_{\rm i}} \left[1-(1-Q_{k'_{\rm i}k'_{\rm o},*})^{k'_{\rm o}}\right]$	$\sum_{k'_{\rm i},k'_{\rm o}} \mathbf{Pr}(k'_{\rm i},k'_{\rm o}) \left[1-(1-Q_{k'_{\rm i}k'_{\rm o},*})^{k'_{\rm o}}\right]$
VI. Mixed Directed and Undirected, Correlated	$Q_{\vec{k}u} = \sum_{\vec{k}} P^{(\rm u)}(\vec{k}'\vert\,\vec{k}) B_{**\vec{k}'} \left[1-(1-Q_{\vec{k}'u})^{k'_{\rm u}-1}(1-Q_{\vec{k}'o})^{k'_{\rm o}}\right]$ $Q_{\vec{k}o} = \sum_{\vec{k}'} P^{(\rm i)}(\vec{k}'\vert\,\vec{k}) B_{**\vec{k}'} \left[1-(1-Q_{\vec{k}'u})^{k'_{\rm u}}(1-Q_{\vec{k}'o})^{k'_{\rm o}}\right]$	$\sum_{\vec{k}'} \mathbf{Pr}(\vec{k}') \left[1-(1-Q_{\vec{k}'u})^{k'_{\rm u}}(1-Q_{\vec{k}'o})^{k'_{\rm o}}\right]$

TABLE I: For the six classes of random networks described in Sec. V, the probability of triggering a global spreading events due to (1) an infected edge, and (2) an infected, randomly chosen single seed (see Eqs. 2 and 4). We indicate by the symbol $*$ when no node or edge type is relevant.

edge of type λ. Due to the Markovian nature of random networks, this probability must satisfy a (nonlinear) recursion relation:

$$Q_{\nu\lambda} = \sum_{\nu'\in\mathcal{N}} \mathbf{Pr}(\nu\lambda\nu') B_{\nu\lambda\nu'} \left[1 - \prod_{\lambda'\in\Lambda_{\nu'}\setminus\bar{\lambda}} (1-Q_{\nu'\lambda'})\right], \tag{2}$$

an expression which involves three elements. First, we have $\mathbf{Pr}(\nu\lambda\nu')$ which is the probability that the edge λ leads to a node of type ν'. The second term is the $B_{\nu\lambda\nu'}$ which as we have just defined is the probability of successful infection. The last term contains the recursive structure. At least one of the edges leading away from the type ν' node must generate a global spreading event (note that the incident edge of type $\bar{\lambda}$ is excluded in the product). The probability this happens is the complement of the probability that none succeed, $\prod_{\lambda'\in\Lambda_{\nu'}\setminus\bar{\lambda}} (1-Q_{\nu'\lambda'})$. Eq. (2) will rarely be analytically tractable (but see [32] for an exactly solved simple model), and will usually be solved numerically by straightforward iteration.

The probability that an infected type ν node seeds a global spreading event follows as

$$Q_\nu = 1 - \prod_{\lambda\in\Lambda_\nu} (1-Q_{\nu\lambda}), \tag{3}$$

where again success is defined in terms of not failing. Finally, the probability that the sole infection of a randomly chosen node leads to a global spreading event is

$$Q = \sum_{\nu\in\mathcal{N}} \mathbf{Pr}(\nu) Q_\nu = \sum_{\nu\in\mathcal{N}} \mathbf{Pr}(\nu) \left[1 - \prod_{\lambda\in\Lambda_\nu} (1-Q_{\nu\lambda})\right]. \tag{4}$$

The effects of weighted triggering schemes can be examined by replacing $\mathbf{Pr}(\nu)$ with the appropriate distribution.

V. APPLICATION TO A SET OF RANDOM NETWORK FAMILIES

In Tab. I, we list the forms of $Q_{\nu\lambda}$ and Q for six specific families of random networks which we describe below. The last of these network families is the most general and contains the other five as special cases. Nodes potentially have three kinds of unweighted edges incident to them: undirected, in-directed, and out-directed, and we use the vector representation $\vec{k} = [k_{\rm u}, k_{\rm i}, k_{\rm o}]^{\rm T}$ to define node classes [14, 27]. The specific transition probabilities, $P^{(\rm i)}(\vec{k}\vert\vec{k}')$, $P^{(\rm o)}(\vec{k}\vert\vec{k}')$, and $P^{(\rm u)}(\vec{k}\vert\vec{k}')$, give the probabilities of an edge leading from a degree $\vec{k}'$ node to a degree $\vec{k}$ node being oriented as undirected, incoming, or outgoing (see Refs. [27] and [32] for more details). For uncorrelated networks, we use the notation $P^{(\mathrm{i})}(\vec{k}|*)$, etc. Similarly for the triggering probabilities, where the node or edge type is irrelevant we also use $*$ (e.g., Q_{**} instead of $Q_{\nu\lambda}$ for undirected, uncorrelated, unweighted networks). For simplicity, we assume infection is due only to properties of the node potentially being infected, which for these networks means the node's degree.

VI. CONCLUDING REMARKS

We have shown that the probability of a single infected node generating a global spreading event can be derived in a straightforward way for spreading processes on a very general class of correlated random networks. Our approach brings a physical intuition to the problem, and while more sophisticated mathematical analyses arrive at the same results, and are certainly useful for more detailed investigations, they are burdened with some degree of inscrutability.

Acknowledgments

PSD was supported by NSF CAREER Award # 0846668; KDH was supported by VT-NASA EPSCoR; JLP was supported by NIH grant # K25-CA134286. The authors are grateful for the computational resources provided by the Vermont Advanced Computing Center which is supported by NASA (NNX 08A096G).

[1] P. J. Richerson and R. Boyd, *Not by Genes Alone* (University of Chicago Press, Chicago, IL, 2005).
[2] A. Chmiel, J. Sienkiewicz, M. Thelwall, G. Paltoglou, K. Buckley, A. Kappas, and J. A. Hoyst, PLoS ONE **6**, e22207 (2011).
[3] D. M. Romero, B. Meeder, and J. Kleinberg, in *Proceedings of World Wide Web Conference* (2011).
[4] P. Rozin and E. Royzman, Personality and Social Psychology Review **5**, 296 (2001).
[5] J. Leskovec, L. Backstrom, and J. Kleinberg, in *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (2009), pp. 497–506.
[6] J. Berger and G. Le Mens, Proc. Natl. Acad. Sci. **106**, 8146 (2009).
[7] A. V. Banerjee, Quart. J. Econ. **107**, 797 (1992).
[8] S. Bikhchandani, D. Hirshleifer, and I. Welch, J. Polit. Econ. **100**, 992 (1992).
[9] E. Rogers, *The Diffusion of Innovations* (Free Press, New York, 1995), Fifth ed.
[10] S. G. Barsade, Administrative Science Quarterly **47**, 644 (2002).
[11] P. Sieczka, D. Sornette, and J. A. Holyst, Eur. Phys. J. B **82**, 257 (2011).
[12] A. D. Cliff, P. Haggett, J. K. Ord, and G. R. Versey, *Spatial diffusion: an historical geography of epidemics in an island community* (Cambridge University Press, Cambridge, UK, 1981).
[13] V. Colizza, A. Barrat, M. Barthelmey, A.-J. Valleron, and A. Vespignani, PLoS Medicine **4**, e13 (2011).
[14] M. Boguñá and M. Ángeles Serrano, Phys. Rev. E **72**, 016106 (2005).
[15] J. P. Gleeson and D. J. Cahalane, Phys. Rev. E **75**, 056103 (2007).
[16] J. P. Gleeson, Phys. Rev. E **77**, 046117 (2008).
[17] L. A. Meyers, M. Newman, and B. Pourbohloul, J. Theor. Biol. **240**, 400 (2006).
[18] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[19] D. J. Watts, Proc. Natl. Acad. Sci. **99**, 5766 (2002).
[20] J. P. Gleeson, S. Melnik, and A. Hackett, Phys. Rev. E **81**, 066114 (2010).
[21] A. Hackett, S. Melnik, and J. P. Gleeson, Phys. Rev. E **83**, 056107 (2011).
[22] Y. Ikeda, T. Hasegawa, and K. Nemoto, Journal of Physics: Conference Series **221**, 012005 (2010).
[23] P. Munz, I. Hudea, J. Imad, and R. J. Smith?, in *Infectious Disease Modelling Research Progress*, edited by J. M. Tchuenche and C. Chiyaka (Nova Science Publishers, Inc., 2009), pp. 133–150.
[24] D. J. Watts and P. S. Dodds, in *The Oxford Handbook of Analytical Sociology*, edited by P. Hedström and P. Bearman (Oxford University Press, Oxford, UK, 2009), chap. 20, pp. 475–497.
[25] H. S. Wilf, *Generatingfunctionology* (A K Peters, Natick, MA, 2006), 3rd ed.
[26] M. E. J. Newman, SIAM Review **45**, 167 (2003).
[27] P. S. Dodds, K. D. Harris, and J. L. Payne, Phys. Rev. E **83**, 056122 (2011).
[28] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, Comput. Netw. **33**, 309 (2000).
[29] T. C. Schelling, J. Conflict Resolut. **17**, 381 (1973).
[30] M. Granovetter, Am. J. Sociol. **83**, 1420 (1978).
[31] J. D. Murray, *Mathematical Biology* (Springer, New York, 2002), Third ed.
[32] J. L. Payne, K. D. Harris, and P. S. Dodds, Phys. Rev. E **84**, 016110 (2011).