On the Properties of Infective Flooding in Low-Duty-Cycle Networks

Luca Baldesi, Leonardo Maccari, Renato Lo Cigno

Dept. of Information Engineering and Computer Science, University of Trento, Italy

{luca.baldesi,leonardo.maccari,renato.locigno}@unitn.it

Abstract—Broadcasting information in a network is an important function in networking applications. In some networks, as wireless sensor networks or some ad-hoc networks it is so essential as to dominate the performance of the entire system. Exploiting some recent results based on the computation of the eigenvector centrality of nodes in the network graph and classical dynamic diffusion models on graphs, this paper derives a novel theoretical framework for efficient information broadcasting in mesh networks with low duty-cycling without the need to build a distribution tree. The model provides lower and upper stochastic bounds with high probability. We show that the lower bound is very close to the theoretical optimum and that a preliminary implementation provides results that are very close to the lower bound on classical graph models.

I. INTRODUCTION

Broadcast, the function of sending a piece of information to all nodes in a network, is a fundamental and pervasive function in many protocols, applications, and network architectures as well. Flooding of Link-State (LS) advertisements in LS routing protocols or streaming in a multicast group with Peer-to-Peer (P2P) technologies are examples of broadcast in application overlays. In wireless ad-hoc networks such as Wireless Sensor Networks (WSNs) it is normally executed on the physical topology (as opposed to a logical overlay) and it is so important that its performance impacts the overall network efficiency. In these networks, which are often considered the base of Internet of Things (IoT), broadcasting pertains to sensor data, queries, or messages about diagnosis, localization, routing, and configuration: In practice in every domain of operation [1].

If in traditional networks and overlays broadcasting can be often solved satisfactorily with techniques that builds a distribution tree [2], [3] or with brute-force approaches such as limited flooding (as in Open Shortest-Path First (OSPF) LS advertisements), in WSN there are three additional challenges: i) Dynamism; ii) Energy consumption; and iii) Duty cycling, i.e., the ratio between the wake-up time of the node and the overall time of the cycle [4] that can be as low as 0.01 or even less [5]. A WSN is dynamic, meaning that even if the nodes are stationary, the surrounding conditions vary and at every instant we expect a few links to appear or disappear. Energy efficiency is hampered by continuous signaling, reduction in duty cycling, need for overhearing messages. And low duty cycle means that any reconfiguration takes a long time as nodes seldom wake up. For these reason we consider a broadcasting strategy that does not rely on trees, that are intrinsically fragile even in presence of minimal modifications of the topology. In the rest of the paper we concentrate, as reference scenario, on WSNs, albeit our results are general and apply to any network, physical or logical. In this context we use the term *flooding* to identify the broadcast process, while the term *broadcast* is used only to refer to physical layer broadcast, if used.

The contribution of this paper stems from the encounter of a recent result published in the context of P2P streaming [6], with classical results from epidemic diffusion based on differential equations [7], [8]. We look at the broadcasting process as the "infective" propagation of data on a graph, that models the given topology of the network, hence the title, and we modify the infective capability of each node in such a way to optimize the information distribution. The result is a sound theoretical framework providing bounds either for the maximum delay and the energy consumption required for flooding operations. Flooding is performed exploiting all the resources of the network, without building a distribution tree, thus resulting in a very robust system that requires minimal signaling.

The theoretical results we derive are confirmed with simulations on different topologies, while performance comparisons with other techniques such as the tree-based solution proposed in [5] are left for future work.

II. SYSTEM MODEL

We consider a connected, multi-hop network described by an undirected graph $G(V, E)$, where V is the set of nodes and $E \cap (V \times V)$ is the set of edges. The network is stationary or with slow mobility as assumed also in [9], [5], so that in general the network topology does not change too much from one wake-up cycle to the next. One node, called source, is the originator of the packets to be flooded, but we do not make any further assumption on its location in G; it is likely that the source changes from one flooding event to another. This scenario is typical of a WSN used for monitoring, in which at a certain instant a sensor detects a certain event and alerts the other nodes of the network. Similarly flooding is needed for time synchronization of nodes, which can be triggered by any of the sensors acquiring time from an external source [10].

Notice that in tree-based flooding if the source changes the flooding tree must be recomputed, e.g., using Djikstra algorithm, unless a global Minimum Spanning Tree is used, which however does not minimize the distribution delay. Moreover the time interval between two flooding events is
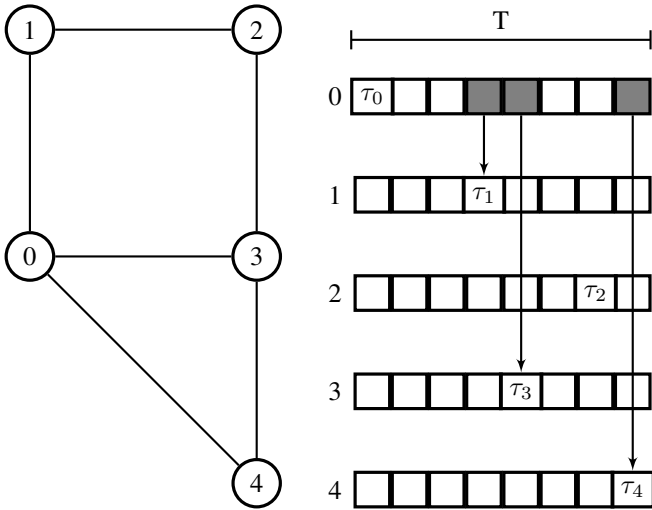
Fig. 1. The network model consists of a mesh network whose nodes have periodic listening periods τ_i and wake up when necessary in period τ_j to transmit to node j; transmissions are unicast as τ_i are separated; in the picture node 0 sends a packet (in gray) to nodes 2, 3, and 4.

orders of magnitude larger than the wake-up cycle, so changes in the topology are likely between two flooding events. Tree structures are fragile and must be maintained over time, requiring a very large signaling overhead for low duty-cycle networks.

Let T be the wake-up cycle that is divided in fixed lengths time slots of τ seconds. τ is long enough to synchronize a transmitter and a receiver and to transmit a packet. For the sake of simplicity we assume that a transmission is successful, as our focus is on the assignment of transmission resources to nodes based on their topological position to minimize the flooding delay, so re-transmissions are not essential to the problem, and can be solved with a simple ARQ protocol, which is left for future work. Each node wakes up for a single τ slot during a cycle T to listen for incoming packets and sleeps during the others to reduce energy consumption, or for any other reason. This assumption is the same adopted, for instance, in [9]. We assume there is an initialization phase (how it is done is outside the scope of this paper, see for instance the solution proposed in [11]) during which nodes select their listening periods and exchange them with their neighbors. After the initialization, nodes wake up if they have to send packets during other nodes listening periods or if they have a scheduled listening period.

Figure 1 depict the situation with 4 nodes and a simple topology. τ_i identifies the listening slot of duration τ selected by node i. As the duty-cycle is very low, one can assume that waking periods are different in the entire network, minimizing interference, but it is enough that they are different in 2-hop neighborhoods.

In the following subsections we combine our results on the optimality of receiver-equal resource allocation [6] together with an infective model dissemination on graphs to derive

bounds on the performance of information flooding on mesh networks with minimal signaling.

A receiver-equal strategy guarantees that in a streaming application, in which a source injects into the network a sequence of packets, every node receives the same amount of information at steady state given that the overall resources are minimal. Minimal resources means that, if B_s bit/s is the stream bandwidth, then the overall capacity allocated in the network is $|V| \times B_s$. This resource allocation to every node is proportional to the eigenvector centrality of the node in $G(V, E)$[1].

An infective model is, in some sense, the mathematical formulation of opportunistic (or stochastic) flooding. Each node that possesses the information pass it to one of its neighbors following some protocol, until all his neighbors have received the information. When this is true for all nodes, then the information has been flooded to the entire network.

The goal of this paper is to show that the steady-state reception-equal property can be applied to flooding of one packet and derive time bounds for the average case.

A. Reception-Equal P2P Streaming

In a P2P (Peer-to-Peer) streaming application there is a source of the stream that injects packets in the network. Every node (including the source) at every instant has a buffer of packets that can share with its neighbours, and decides which packet to share with which neighbour randomly. As stated above, at steady state, the optimal resource allocation is the one that guarantees that every node receives the same amount of information (the same number of packets) per time-interval. The result presented in [6] can be summarized as follows; Section III derives delay bounds starting from this result.

Let A' be a stochastic transition matrix for G so that the element $A'_{ij} \in [0, 1]$, $A'_{ij} > 0 \iff (i, j) \in E$ represents the probability for node j to send a packet to node i during a cycle T and $\vec{1}^T A' = \vec{1}^T$. Let Θ_j be the throughput (in terms of packets per second) that node j sustains in average. Then, from the Theorem 1 in [6],

$$\Theta_j = \alpha \sum_{k=1}^{|V|} \frac{A'_{kj} x_j}{x_k}, \quad A_{ij} = \frac{\frac{A'_{ij}}{x_i}}{\sum_{k=1}^{|V|} \frac{A'_{kj}}{x_k}} \tag{1}$$

are such that:

$$\alpha \vec{1} = A\Theta \tag{2}$$
$$|\Theta| = \alpha|\vec{1}| = \alpha|V| \tag{3}$$
$$\vec{1}^T A = \vec{1}^T \tag{4}$$
$$A'_{ij} = 0 \iff A_{ij} = 0 \tag{5}$$

where $x_i \in \mathbb{R}$ is the eigenvector centrality of node i and $\alpha \in \mathbb{R}^+$ is an amplification parameter. The theorem states that the new stochastic transition matrix A describes the same links as A' but with different values (Eqs. (4) and (5)). Θ_j represents

[1]The definition and properties of the eigenvector centrality are derived from [12]

the number of packets node j sends during T and, overall, this number is, on average, α per node (Eq. (3)). In this work we set $\alpha = 1$ as increasing it means using more resources (bandwidth and energy), which we are instead interested in minimizing. Equation (2) grants that every node has the same probability of receiving the information if we average over all possible sources $s \in V$.

If A' is column-uniform, which means that pakets are sent with uniform probability to the neighbors, these parameters can be computed by the nodes distributively by simply gossiping their neighbourhood set size [6].

B. The Infective Model

The flooding of a packet in a network can be seen as a virus propagation starting at the source node, and all nodes being susceptible to the infection while they do not have the packet and infective when they have it. We are interested in studying and characterizing the speed of such infection.

Our infection process corresponds to the elementary SI model: a node can be in either one of the two states, susceptible (S) or infected (I), there is no recovery from the infection and nodes remain infectious indefinitely (they do not die or recover from the infection) [7]. We are aware that there is a large body of literature on disease spreading, obviously in the medical literature, but also in networking (see for instance [13], [8], [14], [15], [16] and references in these works), but indeed the simple SI model represents exactly what happens flooding a packet into a network, taking into account the topological properties of the network graph G.

The initial spread of a virus in a network subject to the SI model is exponentially fast [7]. The speed depends on the largest eigenvalue of A' (1 in our case as it is column stochastic) and the rate of infection. During this initial phase, the nodes with large eigenvector centralities are more likely to be infected [7].

We denote with $y_i(k)$ the probability of node i to be infected at time k (we use discrete time to better map with time cycle T), with $S(k)$, $I(k)$ the group of susceptible and infected nodes at time k, and N_i is the set of neighbour nodes of i. Hence, the following dynamic equation holds:

$$y_i(k+1) = y_i(k) + P\{i \in S(k), j \in I(k), j \text{ infects } i, \forall j \in N_i\} \tag{6}$$

Equation (6) states the probability of node i to be infected at time $k+1$ is given by the same probability at the previous time step plus the probability of transition from the susceptible state to the infected one, which occurs if at least one neighbour j is infected (at time k) and pass the infection.

Unfortunately, Eq. (6) cannot be integrated in closed form conditioned on the graph topology, and it is hence difficult to handle mathematically. To ease the analysis we take advantage from its first order approximation (see Section V for a discussion on this approximation):

$$y_i(k+1) = y_i(k) + (1 - y_i(k))(1 - P\{j \in I(k), \forall j \in N_i, \ j \text{ does not infect } i\}) \tag{7}$$

In the case of packets flooding using the reception-equal strategy we have that j infects i (j sends a packet to i) with probability $A_{ij}\Theta_j$ (the throughput of j multiplied by the probability of sending a packet to neighbor i). Equation (7) can be expressed in closed form as:

$$y_i(k+1) = y_i(k)+$$
$$[1 - y_i(k)] \left[1 - \prod_{j=1}^{|V|} [1 - y_j(k) A_{ij} \Theta_j] \right] \tag{8}$$

the next section exploits this approximation to derive closed form upper and lower delay and energy bounds.

III. Flooding Delay Bounds

In this section we present our stochastic upper and lower bounds for the distribution delay time. To this end we assume an initial uniform probability that each node in a network can be the source, i.e., $y_i(0) = \frac{1}{|V|} \ \forall i$.

We consider the following difference equation which we call the upper bound:

$$\begin{cases} \omega(k+1) &= 2\omega(k) - \omega^2(k) \\ \omega(0) &= \frac{1}{|V|} \end{cases} \tag{9}$$

And this other difference equation, referred to as the lower bound:

$$\begin{cases} \Omega(k+1) &= 2\Omega(k) - \frac{3}{2}\Omega^2(k) + \frac{\Omega^3(k)}{2} \\ \Omega(0) &= \frac{1}{|V|} \end{cases} \tag{10}$$

Note that both Eqs. (9) and (10) have two fixed points $\{0, 1\}$ the latter of which is attractive. Hence, $\omega(k), \Omega(k)$ are monotonically increasing functions and, given their initial value $\omega(0) = \Omega(0) = \frac{1}{|V|}$ their values are in the interval $[\frac{1}{|V|}, 1)$.

We are now ready to state the main finding of this work,

Theorem 1. *Let nodes have an equal initial probability* $y_i(0) = \frac{1}{|V|}$, $\forall i$ *and* $\alpha = 1$, *than we have:*

$$\Omega(k) \leq y_i(k) \leq \omega(k), \ \forall i, k$$

Proof. First note that given the reception-equal property (Eq. (2)) the following identities holds for any i:

$$\omega(z+1) = 2\omega(z) - \omega^2(z) =$$
$$\omega(z) + (1 - \omega(z)) \sum_{j=1}^{|V|} \omega(z) A_{ij} \Theta_j =$$
$$\omega(k) + (1 - \omega(k)) \left[1 - \left(1 - \sum_{j=1}^{|V|} \omega(k) A_{ij} \Theta_j \right) \right] \tag{11}$$

and

$$\Omega(k+1) = 2\Omega(k) - \frac{3}{2}\Omega^2(k) + \frac{\Omega^3(k)}{2} =$$

$$\Omega(k) + (1-\Omega(k))\left[\sum_{j=1}^{|V|}\Omega(k)A_{ij}\Theta_j + \right.$$

$$\left. -\frac{1}{2}\sum_{j=1}^{|V|}\sum_{z=1}^{|V|}\Omega(k)A_{ij}\Theta_j\Omega(k)A_{iz}\Theta_z\right] =$$

$$\Omega(k) + [1-\Omega(k)]\left[1-\left(1-\sum_{j=1}^{|V|}\Omega(k)A_{ij}\Theta_j + \right.\right.$$

$$\left.\left.\frac{1}{2}\sum_{j=1}^{|V|}\sum_{z=1}^{|V|}\Omega(k)A_{ij}\Theta_j\Omega(k)A_{iz}\Theta_z\right)\right]$$

The last element can be represented compactly as $\frac{1}{2}\sum_{j=1}^{|V|}\sum_{z=1}^{|V|}b_jb_z$ where $b_j = \Omega(k)A_{ij}\Theta_j \in \mathbb{R}^+$. Since $\frac{1}{2}\sum_{j=1}^{|V|}\sum_{z=1}^{|V|}b_jb_z = \sum_{j=1}^{|V|}\sum_{z>j}^{|V|}b_jb_z + \frac{1}{2}\sum_{j=1}^{|V|}b_j^2$, then $\frac{1}{2}\sum_{j=1}^{|V|}\sum_{z=1}^{|V|}b_jb_z \geq \sum_{j=1}^{|V|}\sum_{z>j}^{|V|}b_jb_z$ and,

$$\Omega(k+1) \leq \Omega(k)+$$

$$[1-\Omega(k)]\left[1-\left(1-\sum_{j=1}^{|V|}\Omega(k)A_{ij}\Theta_j + \right.\right.$$

$$\left.\left.\sum_{j=1}^{|V|}\sum_{z>j}^{|V|}\Omega(k)A_{ij}\Theta_j\Omega(k)A_{iz}\Theta_z\right)\right] \qquad (12)$$

Let's first prove $y_i(k) \leq \omega(k)$ by induction over k,
k=0) $\omega(0) \geq y_i(0)$ $\forall i$ by definition,
k=z+1) Let's assume $\omega(k) \geq y_i(k)$ $\forall i, k = 1, \ldots, z$

$$\omega(z+1) \geq \omega(z) + (1-\omega(z))\left[1-\prod_{j=1}^{|V|}(1-\omega(z)A_{ij}\Theta_j)\right]$$

Given Eq. (2) and $\alpha = 1$ we have that $A_{ij}\Theta_j \leq 1$ $\forall i,j$, then

$$y_i(k)A_{ij}\Theta_j < 1, \omega(k)A_{ij}\Theta_j < 1, \Omega(k)A_{ij}\Theta_j < 1$$

and thus we can apply Proposition 1 (considering $a_j = \omega(z)A_{ij}\Theta_j$). For simplicity we call

$$\Gamma = 1 - \prod_{j=1}^{|V|}(1-\omega(z)A_{ij}\Theta_j)$$

then,

$$y_i(z+1) = y_i(z) + (1-y_i(z))\left[1-\prod_{j=1}^{|V|}(1-y_j(z)A_{ij}\Theta_j)\right]$$
$$\leq y_i(z) + (1-y_i(z))\Gamma$$

as, because of the inductive step, $\omega(z) \geq y_i(z)$ $\forall i, z$.

$$\omega(z+1) - y_i(z+1) = \omega(z) - y_i(z) + (y_i(z) - \omega(z))\Gamma =$$
$$(\omega(z) - y_i(z))(1-\Gamma)$$

as $\omega(z) \geq y_i(z)$ and $\Gamma < 1$ then $(\omega(z) - y_i(z))(1-\Gamma) \geq 0$ and $\omega(z+1) \geq y_i(z+1)$.

The proof of $y_i(k) \geq \Omega(k)$ is again by induction over k,
k=0) $\Omega(0) \leq y_i(0)$ $\forall i$ by definition,
k=z+1) Let's assume $\Omega(k) \leq y_i(k)$ $\forall i, k = 1, \ldots, z$

$$\Omega(z+1) \leq \Omega(z) + (1-\Omega(z))\left[1-\prod_{j=1}^{|V|}(1-\Omega(z)A_{ij}\Theta_j)\right]$$

combining Eq. (12) and Proposition 2 (considering $a_j = \Omega(z)A_{ij}\Theta_j$). For simplicity we call

$$\gamma = 1 - \prod_{j=1}^{|V|}(1-\Omega(z)A_{ij}\Theta_j)$$

$$y_i(z+1) \geq y_i(z) + (1-y_i(z))\gamma$$

as, because of the inductive step, $\Omega(z) \leq y_i(z)$ $\forall i, z$.

$$y_i(z+1) - \Omega(z+1) = (y_i(z) - \Omega(z))(1-\gamma)$$

as $\Omega(z) \leq y_i(z)$ and $\gamma < 1$ then $(y_i(z) - \Omega(z))(1-\gamma) \geq 0$ and $\Omega(z+1) \leq y_i(z+1)$. $\square$

Theorem 1 exploits the first order approximation of the SI model on a graph G given by Eq. (8), with the reception-equal property granted by Eq. (2) to derive theoretical stochastic upper and lower bounds for the probability that node i is infected, i.e., it has received the information, at time k. To ensure that these bounds hold with minimal resource use we set $\alpha = 1$; proofs of the inequalities used are reported in Appendix A.

A node-independent bound express the probability that a generic node has received the packet regardless of its position in the network averaged on all the possible sources of the information. These bounds can also be interpreted as bounds on the information delay expectation for each node when there is no knowledge on the information source position, or in the SI terminology, when the initial probability of infection is $y_i(0) = \frac{1}{|V|}\forall i$.

A. Solving the bounds

Equation (9) is a second order difference equation similar to the *logistic map*, but its parameters keep it in the stability region (it is not chaotic), furthermore we take advantage that we are only interested in studying its value for $\omega(k) \in [0,1]$. Equation (9) has two fixed points, $\omega(k) = \{0,1\}$. The first one is irrelevant as $\omega(0) > 0$ and the latter is an attractor as Eq. (9) is non-decreasing.

Let $\omega(k) = 1 - \epsilon_k$ be the probability that node i is infected at time k, with $\epsilon_0 = 1 - \frac{1}{|V|}$, then we have

$$\omega(k+1) = 1 - \epsilon_{k+1} = 2(1-\epsilon_k) - (1-\epsilon_k)^2 =$$
$$2 - 2\epsilon_k - 1 + 2\epsilon_k - \epsilon_k^2 = 1 - \epsilon_k^2$$

and consequently $\epsilon_{k+1} = \epsilon_k^2$ that finally implies

$$\omega(k) = 1 - \epsilon_0^{2^k} = 1 - \left(1 - \frac{1}{|V|}\right)^{2^k} \qquad (13)$$

and solving for k

$$k = \left\lceil \log_2 \left(\log_{(1 - \frac{1}{|V|})} (1 - p) \right) \right\rceil, \; \forall \, p \in \left(\frac{1}{|V|}, 1 \right) \quad (14)$$

where k is the average umber of time cycles needed for a node to have received a packet with probability p.

Equation (13) indicates that the reception-equal condition grants, regardless of the network topology of G, a double exponential speed of convergence (much faster than the exponential speed in the SI model) in the early distribution phase (when $y_i(k) \ll 1, \; \forall i$).

Eq. (10) is strictly non-decreasing for $\Omega(k) \in [0, 1)$ and with a slower growth than Eq. (9). Unfortunately, Eq. (10) cannot be stated in closed form but we can numerically integrate the difference equation.

B. Energy consumption

The time at which each node has received the packet with probability p is at most $k = \Omega^{-1}(p)$ when averaged on all sources. For Eq. (3) the total number of packets sent during the flooding up to time k is $\alpha k |V|$. From a previous result [6], we also know that each peer j sends $k\Theta_j$ during k cycles.

Note that, it is also granted that $\Theta_j \leq \alpha |N_j|$ where N_j is the neighbour set of node j. Such bound grants node energy consumption to be contained. If a node is low on battery and does not want to contribute to the flooding, it can simply reduce its neighbour set (dropping some links, or simply avoiding to communicate their presence) and let the system recompute the optimal parameters.

Finally, it is worth noticing we are assuming minimal signalling among the nodes which may end up sending duplicated content.

IV. Numerical Results

To further demonstrate the applicability of our theoretical results we simulate with a simple SI model script the spreading of the packet in networks. We compare with $\omega^{-1}(p)$ and $\Omega^{-1}(p)$ the confidence interval of the value of slot k at which all the nodes have received the packets for several runs on the same network using random source node with for $p = 0.9999$ (assumed as certainty of reception).

We perform tests using Barabási-Albert and Erdős-Rényi networks for their well-known properties. Test networks target a constant density for the ease of comparison, $\frac{|E|}{|V|} \sim 4$. For each network, the flooding is simulated 100 times, picking a different node as the source.

We indicate with *RE* the reception-equal optimized system of Section II-A, with *SE* the sending-equal standard system without the optimization and with *RE,opt* the reception-equal strategy with trivial distribution optimization.

First, we are interested to verify our bounds are correctly identifying the reception delay of the flooding. Figs. 2 and 3 report the values of $\omega^{-1}(p), \Omega^{-1}(p)$ for networks with $|V| \in \{100, 300, 500, 750, 1000\}$. As can be seen our bounds characterize the distribution of the packet reception delay in the whole network (indicated as *Delay (RE)*), as the delay
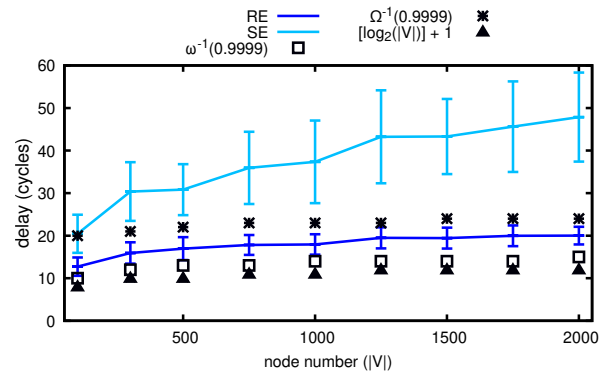

Fig. 2. Mean and standard deviation for reception delay and thresholds for Barabási-Albert networks varying $|V|$.
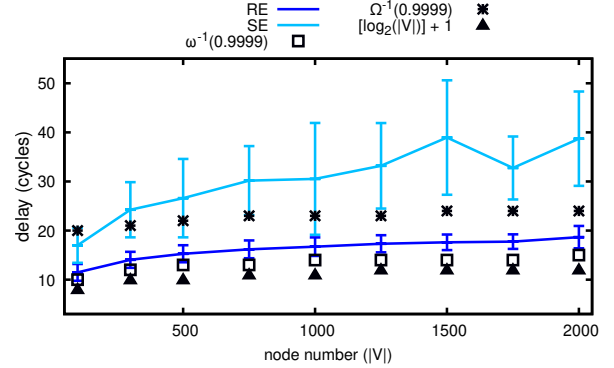

Fig. 3. Mean and standard deviation for reception delay and thresholds for Erdős-Rényi networks varying $|V|$.

confidence interval correctly falls in between the bounds. As we can see varying $|V|$, such characterization is consistent varying the network size.

For the sake of comparison, Figs. 2 and 3 also show the delay distribution on the same networks without the optimization described in Section II-A, (indicated as *Delay*). As it can be seen, while the reception-equal strategy grants the delay to grow slowly in between the bounds, the naïve one diverges
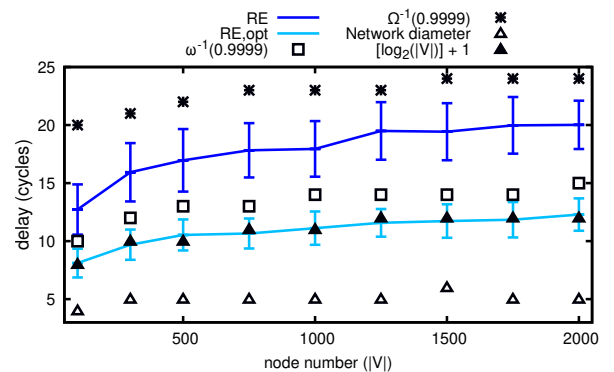

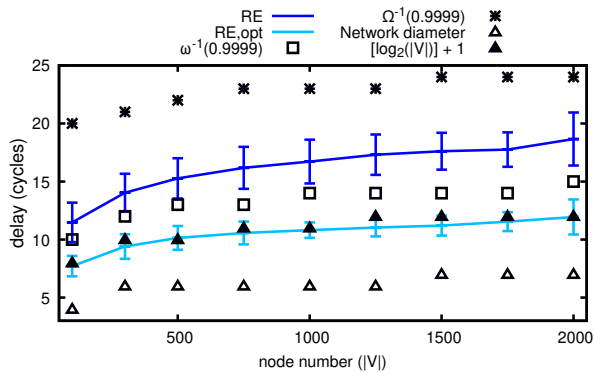Fig. 4. Mean and standard deviation for reception delay and thresholds for Barabási-Albert networks varying $|V|$.

Fig. 5. Mean and standard deviation for reception delay and thresholds for Erdős-Rényi networks varying $|V|$.

steadily. Also the confidence interval deviation is sensibly smaller for the optimized case.

Figs. 2 and 3 show also an important reference bound [**?**], dictating the lowest possible delay in a full-mesh Peer-to-Peer (P2P) network with complete knowledge of node packet reception (no duplicate packet transmission) and single transmission for time slot (each node can send only one packet per cycle T). Such bound is important as we can compare with a full-knowledge bound for the same flooding process in a mesh network, though without any distribution optimization. The reception-equal distribution delay, Delay (RE), is close to this bound and, with little surprise, does not violate it. In fact, the P2P lower bound is indeed a lower bound for $\omega^{-1}(0.9999)$.

While results shown so far, are related to the simulation of Eq. (6) for the ease of comparison with the bounds Ω, ω, we can go a step forward and design a more intelligent distribution. Toward an actual useful implementation, a simple and straightforward optimization is to prevent nodes to send the packet to the same destination twice. We call this strategy reception-equal optimized (RE,opt) and Figs. 4 and 5 report its performance in comparison with the previous results. As can be noted the distribution delay achieves the P2P full-knowledge bound performance both in the Barabási-Albert and Erdős-Rényi networks.

To introduce a comparison, although still theoretical, with the flooding strategies exploiting trees, we report in Figs. 4 and 5 the diameter of networks. The network diameter relates to the shortest tree height feasible on the network. We can look at it as the reception delay lower bound for tree distribution-based systems. In Figs. 4 and 5, the performance or our simple optimization are very close to the graph diameter. Further comparison of even more optimized versions of our approach with real tree-based strategies are not in the scope of this paper but, as they seem promising, are left to future works.

V. MODEL LIMITATIONS

While Section IV confirms that our model accurately describes the broadcast function on two widely-used random network models, we have to note that it is not truly universal. Consider for instance a linear network, in which every

node has exactly two neighbors (excluding the nodes at the extremes) and the diameter of the network is $|V|$. In such network at each cycle at most two nodes gets infected. In this corner case the delay needed to achieve $p = 0.9999\%$ grows linearly with $|V|$ and breaks the theoretical bounds we formulated.

The reason for this is to be found in the step from Eq. (6) to Eq. (7) which introduces an implicit approximation. Equation (7) in fact assumes that $P\{i \in S\}$ is independent of $P\{j \in I\}$ and multiplies the two probabilities. In a fully meshed network this assumption is true as all nodes are neighbours so the probability of j to be infected at time k depends only on the total number of infected nodes at time $k-1$ and not on their position. In general this is not true, as $P\{i \in S\}$ strongly depends on $P\{j \in S\}$ if j and i are neighbours, while the inference between $P\{j \in S\}$ and $P\{i \in S\}$ decreases with the distance from j to i. In terms of density and path diversity the linear network is at the opposite extreme of a full mesh, thus it is not surprising that our model fails to capture its behaviour.

Comforted by the positive results we obtained in our numeric evaluations on well known topologies, we leave to future works the sensitivity analysis of our approximation, and the potential adoption of approximations of higher order than the first that are available in the literature [17].

VI. RELATED WORKS

Low-duty-cycle WSNs are said *synchronous* if the node active state happens at a centrally fixed time or *asynchronous* if they are scheduled independently. Literature focuses mostly on asynchronous low-duty-cycle WSNs, where often tree-based dissemination of collection overlays are built. This overview of literature focuses on the one hand on recent papers that address the problem of information flooding (not collection to a sink) in WSNs, and on the other hand on papers that analyze or propose epidemic dissemination techniques, in this case not limited to WSNs, but ranging also from P2P networks to classical infection models used in, or derived from, medical literature.

Wang and Liu [4] proposes a reinterpretation of broadcasting for the context of WSNs and they provide a centralized optimization model from which deriving an approximated distributed solution. Guo et al. [5] address both the delay and the energy constraints deriving a tree-based distribution solution considering lossy links. Cheng et al. [9] propose a flooding tree construction algorithm optimizing with respect to the energy consumption but considering delay bounding. Such algorithm is an approximated distributed version of a centralized optimal one. The work by Niu et al. [18] follows the same scheme as they propose an heuristic algorithm derived from a minimum spanning tree centralized model.

There are also works optimizing existing flooding solutions; Cheng et al. [19] propose the Dynamic Switching-based Reliable Flooding (DSRF) to enhance the reliability of flooding. The flooding optimization by Guo et al. [20] synchronize the active state of nodes sharing the same tree parent node.

Physical channel overhearing has been investigated by Xu et al. [21] as a mean to save delay during message flooding.

Asynchronous Duty-cycle Broadcasting (ADB) is a protocol implemented directly in the MAC layer of WSN nodes which allows flooding by exploiting MAC-layer information.

In contrast to recent publications on flooding on WSN, our approach is fully de-centralized and works with unstructured mesh networks without the aid of trees. That grants an higher degree of robustness against node failure, a lower signalling overhead and promising applications in time-varying networks. Epidemics, the field about modeling and analysing the dynamics of virus spreading, has been prolific in the past decades, though, only recently proper insights on how to control it have been provided [13]. A large part of computer science literature on epidemics focus on malware spreading [22], [15], [14]. Chen et al. [15] use the SIR (Susceptible-Infected-Recovered) model to control dissemination of information in heterogeneous, time-varying networks. The work by Dadlani et al. [14] uses a SIS model instead and provide infection stability results. This work, together with the one by Ganesh et al. [23] highlight the importance of being dependent on a specific network topology for studying epidemics. That is a crucial observation that our approach overcomes exploiting the reception-equal property obtained with the re-assignment of resources based on the eigenvector centrality.

Works by Liu and Buss [24] and by Ogura and Preciado [25] use the SIS model for data dissemination; the former optimizing the node transmission rate while the latter defining exponential growth conditions for time-varying networks.

Other papers deal with different aspects of data dissemination through epidemics approaches. The paper by Chen et al. [16] focuses on delivery dynamics on WSN with cognitive radios, the work by Ramanathan et al. [26] optimize the loss rate for Delay Tolerant Network (DTN) and Byun and So [27] address the context of duty-cycled Wireless Sensor-Actuator Network (WSAN) and propose a scheme to adjust the node transmission rates for user-given delay constraint.

None of these papers, in part also because of their application fields, give delays bounds on information flooding that are independent from the network/graph topology. Up to now it was considered that the optimal strategy to flood information to all nodes of a network could not be independent from its topology. The results we present, instead, show that it is possible to exploit the topological properties of the network to de-correlate the optimal flooding strategy from the topology itself. This observation is what enables the general analysis that in this paper leads to the bounds presented in Section III.

VII. Conclusions and Future Works

Flooding information to all the nodes of a network remains an important function in many networks and applications. Many solutions have been proposed and are working satisfactorily in networks from P2Ps overlays to WSNs, but in many cases remains sub-optimal, e.g., because they have a non-marginal overhead to build a distribution tree, or because they are fragile to topology changes. This paper has presented

fundamental delay bounds for epidemic flooding in low duty-cycle networks that exploit the eigenvector centrality of nodes in the network to allocate resources, i.e., how many copies of the information per time-cycle a node must send. These bounds are, thanks to a property of the resource allocation that we have called *reception equal*, independent of the network topology, a results that is, to the best of our knowledge, presented for the first time. Besides, the lower bound on delay converges with double exponential speed, while the upper bound is exponentially fast, thus ensuring that a proper protocol designed on these properties will converge exponentially fast.

Theoretical bounds on complex graph structures are in general very difficult to derive, while these bounds are not only valid for complex networks, but they are even independent from the network structure. Furthermore, the results are constructive, i.e., they indicate a path to realize a protocol that obtains a performance within the bounds. Indeed, as we discussed, simple heuristics toward a protocol design indicate that actual performance can be very close to the lower bound and in principle even better, because it can exploit additional knowledge that was not included in the theoretical model for mathematical tractability.

Future work on this study starts form the design of a proper protocol that implements the ideas presented in this paper and its comparison with state-of-the art protocols based on trees or other dissemination structures and prosecute with the performance analysis, both theoretical (if possible) and via simulations, of the protocol on networks with different characteristics,.

References

[1] Z. Li, M. Li, J. Liu, and S. Tang, "Understanding the flooding in low-duty-cycle wireless sensor networks," in *IEEE International Conference on Parallel Processing (ICPP)*, 2011.

[2] H.-C. Hsiao and C.-P. He, "A Tree-Based Peer-to-Peer Network with Quality Guarantees," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, pp. 1099 –1110, Aug. 2008.

[3] D. Carra, R. Lo Cigno, and E. W. Biersack, "Stochastic Graph Processes for Performance Evaluation of Content Delivery Applications in Overlay Networks," *IEEE Trans. on Paralled and Distributed Systems*, vol. 19, pp. 247–261, Feb. 2008.

[4] F. Wang and J. Liu, "On reliable broadcast in low duty-cycle wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 5, pp. 767–779, 2012.

[5] S. Guo, L. He, Y. Gu, B. Jiang, and T. He, "Opportunistic flooding in low-duty-cycle wireless sensor networks with unreliable links," *IEEE Transactions on Computers*, vol. 63, no. 11, pp. 2787–2802, 2014.

[6] L. Baldesi, L. Maccari, and R. Lo Cigno, "On the Use of Eigenvector Centrality for Cooperative Streaming," *IEEE Communications Letters*, vol. 21, no. 9, pp. 1953–1956, June 2017.

[7] N. Mark, *Networks: an introduction.* Oxford University Press New York, 2010.

[8] M. Ajelli, R. Lo Cigno, and A. Montresor, "Modeling Botnets and Epidemic Malware," in *IEEE Int. Conf. on Communications (ICC)*, May 2010.

[9] L. Cheng, J. Niu, C. Luo, L. Shu, L. Kong, Z. Zhao, and Y. Gu, "Towards minimum-delay and energy-efficient flooding in low-duty-cycle wireless sensor networks," *Computer Networks*, vol. 134, pp. 66–77, 2018.

[10] K. S. Yildirim and A. Kantarci, "Time synchronization based on slow-flooding in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 1, pp. 244–253, Jan 2014.

[11] P. Dutta and D. Culler, "Practical Asynchronous Neighbor Discovery and Rendezvous for Mobile Sensing Applications," in *6th ACM Conf. on Embedded Network Sensor Systems (SenSys–08)*, Nov. 2008.

[12] P. Bonacich, "Some unique properties of eigenvector centrality," *Social Networks, Elsevier*, vol. 29, no. 4, pp. 555–564, 2007.

[13] C. Nowzari, V. Preciado, and G. Pappas, "Analysis and Control of Epidemics: A Survey of Spreading Processes on Complex Networks," *IEEE Control Systems Magazine*, vol. 36, no. 1, pp. 26–46, Feb 2016.

[14] A. Dadlani, M. Kumar, K. Kim, and K. Sohraby, "Stability and Immunization Analysis of a Malware Spread Model Over Scale-Free Networks," *IEEE Communications Letters*, vol. 18, no. 11, pp. 1907–1910, Nov 2014.

[15] P. Chen, S. Cheng, and K. Chen, "Optimal Control of Epidemic Information Dissemination Over Networks," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2316–2328, Dec 2014.

[16] P. Chen, S. Cheng, and H. Hsu, "Analysis of Information Delivery Dynamics in Cognitive Sensor Networks Using Epidemic Models," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2333–2342, Aug 2018.

[17] J. Benoit, A. Nunes, and M. T. da Gama, "Pair approximation models for disease spread," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 50, no. 1-2, pp. 177–181, 2006.

[18] J. Niu, L. Cheng, Y. Gu, J. Jun, and Q. Zhang, "Minimum-delay and energy-efficient flooding tree in asynchronous low-duty-cycle wireless sensor networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2013.

[19] L. Cheng, J. Niu, Y. Gu, C. Luo, and T. He, "Achieving efficient reliable flooding in low-duty-cycle wireless sensor networks," *IEEE/ACM Transactions on Networking*, no. 6, pp. 3676–3689, 2016.

[20] S. Guo, S. M. Kim, T. Zhu, Y. Gu, and T. He, "Correlated flooding in low-duty-cycle wireless sensor networks." in *19th IEEE Int. Conf. on Network Protocols (ICNP)*, Oct. 2011.

[21] L. Xu, G. Chen, J. Cao, S. Lin, H. Dai, X. Wu, and F. Wu, "Optimizing energy efficiency for minimum latency broadcast in low-duty-cycle sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 11, no. 4, p. 57, 2015.

[22] M. Garetto, W. Gong, and D. Towsley, "Modeling malware spreading dynamics," in *IEEE International Conference on Computer and Communications (INFOCOM)*, March 2003.

[23] A. Ganesh, L. Massoulie, and D. Towsley, "The effect of network topology on the spread of epidemics," in *IEEE International Conference on Computer and Communications (INFOCOM)*, March 2005.

[24] F. Liu and M. Buss, "Optimal control for information diffusion over heterogeneous networks," in *IEEE Conference on Decision and Control (CDC)*, Dec 2016.

[25] M. Ogura and V. Preciado, "Stability of Spreading Processes over Time-Varying Large-Scale Networks," *IEEE Transactions on Network Science and Engineering*, vol. 3, no. 1, pp. 44–57, Jan 2016.

[26] R. Ramanathan, R. Hansen, P. Basu, R. Rosales-Hain, and R. Krishnan, "Prioritized Epidemic Routing for Opportunistic Networks," in *International MobiSys Workshop on Mobile Opportunistic Networking (MobiOpp'07)*, 2007, pp. 62–66.

[27] B. H. and J. So, "Node Scheduling Control Inspired by Epidemic Theory for Data Dissemination in Wireless Sensor-Actuator Networks With Delay Constraints," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 1794–1807, March 2016.

APPENDIX

For the sake of easy reference we report here some mathematical properties of the finite product of real elements, namely $\prod_{i=1}^{n}(1 - a_i)$ in the form of Propositions. We do not claim to have discovered or proven these properties, which may be available in some mathematical handbook, however, they are not trivial and deserve consideration.

A. Product expansion

This section presents some mathematical properties of the finite product of real elements, namely $\prod_{i=1}^{n}(1 - a_i)$.

Proposition 1. *Let $n \in \mathbb{N}$, $a_i \in [0,1) \forall i = 1, \ldots, n$, then*

$$\prod_{i=1}^{n}(1 - a_i) \geq 1 - \sum_{i=1}^{n} a_i$$

Proof. By induction over n:

1) $1 - a_1 \geq 1 - a_1$

k+1) Let's assume the hypothesis is true for $n = 2, \ldots, k \in \mathbb{N}$;

$$\prod_{i=1}^{k+1}(1 - a_i) \geq 1 - \sum_{i=1}^{k+1} a_i \implies$$

$$(1 - a_{k+1})\prod_{i=1}^{k}(1 - a_i) \geq 1 - \sum_{i=1}^{k} a_i - a_{k+1} \implies$$

$$1 - a_{k+1} \geq \left[\frac{1 - \sum_{i=1}^{k}}{\prod_{i=1}^{k}(1 - a_i)}\right] - \left[\frac{a_{k+1}}{\prod_{i=1}^{k}(1 - a_i)}\right]$$

For the inductive step, $\left[\frac{1 - \sum_{i=1}^{k}}{\prod_{i=1}^{k}(1-a_i)}\right] \leq 1$ and, since $a_i \in [0,1)$ $\forall i$, we have $\left[\frac{a_{k+1}}{\prod_{i=1}^{k}(1-a_i)}\right] \geq a_{k+1}$. Hence, proving the hypothesis.

$\square$

Proposition 2. *Let $n \in \mathbb{N}$, $a_i \in [0,1) \forall i = 1, \ldots, n$, then*

$$\prod_{i=1}^{n}(1 - a_i) \leq 1 - \sum_{i=1}^{n} a_i + \sum_{i=1}^{n}\sum_{j>i} a_i a_j$$

Proof. By induction over n:

1) $1 - a_1 \leq 1 - a_1$

k+1) Let's assume the hypothesis is true for $n = 2, \ldots, k \in \mathbb{N}$;

$$\prod_{i=1}^{k+1}(1 - a_i) \leq 1 - \sum_{i=1}^{k+1} a_i + \sum_{i=1}^{k+1}\sum_{j>i} a_i a_j \implies$$

$$(1 - a_{k+1})\prod_{i=1}^{k}(1 - a_i) \leq$$

$$1 - \sum_{i=1}^{k} a_i - a_{k+1} + \sum_{i=1}^{k}\sum_{j>i} a_i a_j + a_{k+1}\sum_{i=1}^{k} a_i \implies$$

$$1 - a_{k+1} \leq$$

$$\left[\frac{1 - \sum_{i=1}^{k} + \sum_{i=1}^{k}\sum_{j>i} a_i a_j}{\prod_{i=1}^{k}(1 - a_i)}\right] + \left[\frac{a_{k+1}\sum_{i=1}^{k} a_i - a_{k+1}}{\prod_{i=1}^{k}(1 - a_i)}\right]$$

For the inductive step, $\left[\frac{1 - \sum_{i=1}^{k} + \sum_{i=1}^{k}\sum_{j>i} a_i a_j}{\prod_{i=1}^{k}(1-a_i)}\right] \geq 1$. Hence, we prove the hypothesis by showing that:

$$-a_{k+1} \leq \left[\frac{a_{k+1}\sum_{i=1}^{k} a_i - a_{k+1}}{\prod_{i=1}^{k}(1 - a_i)}\right] \implies$$

$$-1 \leq \frac{\sum_{i=1}^{n} a_i - 1}{\prod_{i=1}^{n}(1 - a_i)} \implies$$

$$\prod_{i=1}^{n}(1 - a_i) \geq 1 - \sum_{i=1}^{n} a_i$$

Which is given by Proposition 1.

$\square$