Local heuristics and the emergence of spanning subgraphs in complex networks

Alexandre O. Stauffer, Valmir C. Barbosa*

*Universidade Federal do Rio de Janeiro, Programa de Engenharia de Sistemas e Computação, COPPE, Caixa Postal 68511,
21941-972 Rio de Janeiro - RJ, Brazil*

Abstract

We study the use of local heuristics to determine spanning subgraphs for use in the dissemination of information in complex networks. We introduce two different heuristics and analyze their behavior in giving rise to spanning subgraphs that perform well in terms of allowing every node of the network to be reached, of requiring relatively few messages and small node bandwidth for information dissemination, and also of stretching paths with respect to the underlying network only modestly. We contribute a detailed mathematical analysis of one of the heuristics and provide extensive simulation results on random graphs for both of them. These results indicate that, within certain limits, spanning subgraphs are indeed expected to emerge that perform well in respect to all requirements. We also discuss the spanning subgraphs' inherent resilience to failures and adaptability to topological changes.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Complex networks; Local heuristics; Spanning subgraphs

1. Introduction

Let $G = (N_G, E_G)$ be an undirected graph with $n = |N_G|$ nodes and edges representing bidirectional links for pairwise communication among the nodes. We regard G as standing for some unstructured, real-world network whose nodes have no more information on the overall topology of G than can be inferred from their immediate neighborhoods. Given these characteristics, G can also be seen as belonging to the class of networks that have recently come to be referred to as complex networks [2].

Several of the typical problems that require a distributed solution by the nodes of G frequently involve the need to disseminate a piece of information, call it I, through the nodes of the network that share the same connected component with the node that originally possesses I. We assume that this node is unique with respect to that particular information dissemination and refer to it as the originator.

There is a host of possibilities to solve this basic problem of disseminating information through the nodes of G, but invariably they either depend on the existence of a spanning subgraph of G on whose edges the dissemination is performed, or else they employ straightforward flooding of the network's edges by copies of I. The former alternative is often regarded as substantially more cost-effective in terms of several quantities of interest, but of course it carries

---

* Corresponding author.
E-mail address: valmir@cos.ufrj.br (V.C. Barbosa).

with it the inherent need for the desired spanning subgraph to be initially determined and subsequently maintained if the network undergoes topological changes [4].

Subgraphs of interest in this context include the well-known minimum spanning trees, for which several procedures related to creation and maintenance are available [3,6,11], and include also the more general, so-called spanners, which bring with them well-defined structural requirements related to efficiency indicators, but are on the other hand considerably less well-known [13,17]. But regardless of the particular guise of the spanning subgraphs of G for use in information dissemination, the importance of studying them in detail has in recent years found strong justification from the practical side. Notable examples here include the case in which G is some virtual supergraph of a physical network; in this case, spanning subgraphs of G are needed to function as the so-called overlay networks for end-to-end communication over the underlying physical network [7].

In this paper we focus on one very basic question related to determining a spanning subgraph of a complex network G: how close can we get to obtaining a subgraph of G that can be used to disseminate I through the nodes of the originator's connected component, while at the same time satisfying some basic set of performance requirements, if nodes are only allowed to use local information (i.e., information that can be obtained from no farther than the nodes' immediate neighborhoods)? Important requirements involve the expected number of nodes reached when I is disseminated, the expected number of copies of I that are needed, the expected degree of each node in the subgraph (since it relates closely to how many copies of I a node can concurrently send out given its bandwidth limitations), and also the expected path length from the originator on the subgraph.

Even though of a fundamental nature, this question is admittedly too general for an objective analysis. For this reason, we concentrate on the narrower issue of investigating what happens in terms of the aforementioned performance requirements when a subgraph of G is determined in a fully distributed fashion by the nodes according to the following strictly local rules. Each node is responsible for choosing some of its own neighbors in the subgraph and makes its choices in two subsequent steps: first one node is picked from among the node's neighbors; then, with some nonzero probability that we call α, the node gets to pick a second neighbor (which we allow to be identical to the first one it picked).

If this simple procedure is performed by all the nodes of G, then clearly the resulting subgraph, which we denote by $D = (N_D, E_D)$, has $N_D = N_G$ while E_D contains all the edges (u, v) from G such that u chose v at least once or v chose u. This subgraph is then necessarily a spanning subgraph of G; using it for disseminating I from the originator is simply a question of having the originator send I to all of its neighbors in D, and similarly for all the other nodes when they receive I for the first time. Our four performance requirements now have to be examined in terms of how the connected components of D relate to those of G. In particular, does the connected component of D to which the originator belongs span the entire connected component of G that contains it?

Throughout the paper we use c_u to denote the number of choices made by node u and C_u to denote the set of nodes (some of u's neighbors) that get chosen by u. Clearly, $1 \leqslant |C_u| \leqslant c_u \leqslant 2$ and the set of u's neighbors in D is given by C_u, possibly enlarged by every other node v that is a neighbor of u in G and such that $u \in C_v$.

We call D a dissemination subgraph of G and devote the remainder of the paper to analyzing its properties. Our analysis depends, naturally, on the specific criteria that each node uses when making its two decisions. We consider two possibilities, of which the simplest, referred to as the uniform approach, lets each node make its choices uniformly at random among its neighbors. Our analytical treatment of D's properties in this case is given in Section 2; it is based on regarding G as a random graph and makes use of the principles laid down in [16]. This is our core section, and its results are complemented by the simulation results we present in Section 3 for Poisson-distributed node degrees (the classic Erdős–Rényi model [9]) and also for degrees distributed according to a power law (recently discovered to be approximately representative of relevant real-world networks [10,14,2]).

Even though the analytical treatment we offer in Section 2 is specific to this simplest possibility for a choice criterion, at this point it seems to be as far as we have the means to go. For this reason, and notwithstanding the second possibility's clear superiority in terms of our stated performance requirements (see below), we only treat that possibility by means of simulations, whose results are described in Section 3 along with those for the uniform approach. In any event, our mathematical analysis for the uniform approach is innovative and we believe it may yield interesting insight into the analysis of similar problems on complex networks. It may also ultimately be possible to generalize it to handle the more complicated case of the second possibility of choice criteria.

We refer to the second possibility as the degree-based approach. In this approach, the first choice by a node selects uniformly at random from those of its neighbors that have the highest degree (if this is the case for only one neighbor,

then the first choice degenerates into a deterministic decision). The second choice selects a neighbor randomly in proportion to its degree. The degree-based approach is clearly much less uninformed on the network's topology than the uniform approach. For this reason, it is expected to surpass the uniform approach in terms of the indicators we have informally introduced. That this is indeed the case is apparent from the simulation results we show in Section 3. But, even if we find this to be only expected, we also find it remarkable that such a simple strategy of strictly local nature should support a positive answer to our original question to the extent that it does.

We complement our study of our local, two-choice scheme to approximate a spanning subgraph of G by elaborating on its resilience and adaptability properties. These are crucial in the context of networks that may undergo topological changes and we treat them in Section 4. To finalize, we offer concluding remarks in Section 5.

2. Mathematical analysis

Henceforth we regard G as a random graph having node degrees distributed independently from one another and identically to a random variable K_G. Furthermore, the nodes of G are assumed to be connected to one another at random given their degrees, so the degrees of two adjacent nodes remain independent. Our results in this section target the case of a formally infinite set of nodes, that is, the case in which $n \to \infty$.

Let $P_G(a)$ be the probability that a randomly chosen node of G has degree a. The average degree of G, denoted by Z_G, is $Z_G = \sum_{a=0}^{n-1} a P_G(a)$. Also, given the random nature of node interconnections in G, the probability that some node's neighbor has degree b is equal to the expected fraction of edges incident to degree-b nodes, which is given by

$$\frac{b P_G(b)}{\sum_{a=0}^{n-1} a P_G(a)} = \frac{b P_G(b)}{Z_G}. \tag{1}$$

From [15,8,16], we know how to characterize the existence in G of a large, size-$\Theta(n)$ connected component, commonly known as the giant connected component of G (henceforth denoted by GCC_G). If $\langle K_G^2 \rangle$ denotes the second moment of the random variable K_G, that is, $\langle K_G^2 \rangle = \sum_{a=0}^{n-1} a^2 P_G(a)$, then GCC_G almost surely exists if and only if

$$\frac{\langle K_G^2 \rangle}{Z_G} > 2 \tag{2}$$

or, equivalently,

$$\sum_{a=0}^{n-1} \frac{a P_G(a)}{Z_G} a > 2. \tag{3}$$

Intuitively, for a randomly chosen node u and letting v be one of its neighbors, this means that GCC_G almost surely exists (and then G is said to be above the phase transition that gives rise to GCC_G) if and only if v is expected to have strictly more than one neighbor besides u. Otherwise, all the connected components of G are small, consisting of o(n) nodes (and G is said to be below the phase transition).

Now let D be a dissemination subgraph of G constructed by the uniform approach; it is also a random graph, and we let GCC_D denote its giant connected component. Given a randomly chosen, degree-a node u of G, let $\pi_{\mathrm{r}}^{(a)}(1)$ denote the probability that $|C_u| = 1$. If $a > 0$, then

$$\pi_{\mathrm{r}}^{(a)}(1) = 1 - \alpha + \alpha/a, \tag{4}$$

which reflects the probability that either $c_u = 1$ or $c_u = 2$ but both of u's choices were identical. It follows that the probability that $|C_u| = 2$ is

$$\pi_{\mathrm{r}}^{(a)}(2) = 1 - \pi_{\mathrm{r}}^{(a)}(1) = \alpha(a-1)/a. \tag{5}$$

For $a = 0$, clearly $\pi_{\mathrm{r}}^{(a)}(1) = \pi_{\mathrm{r}}^{(a)}(2) = 0$.

We now pause momentarily to note that, throughout the paper, we employ the mnemonic "r" when referring to randomly chosen nodes as in the preceding paragraph. Furthermore, when considering a node v reached by following one of its incident edges, say (u, v) for some neighbor u of v, we utilize the mnemonics "c," "nc, 1," and "nc, 2" as

references, respectively, to the cases of $v \in C_u$, $v \notin C_u$ with $c_v = 1$, and $v \notin C_u$ with $c_v = 2$. These mnemonics are intended to facilitate the use of the probabilities calculated above (and also the ones we are about to calculate) later in this section.

Let u be a randomly chosen node and v one of its neighbors in G such that $v \in C_u$. Nodes u and v are then neighbors in D. Let b be the degree of v in G. The probability that $|C_v \setminus \{u\}| = 0$ (i.e., v chose u and no other neighbor besides u), which we denote by $\pi_c^{(b)}(0)$, is the probability that all of v's choices resulted in u, that is,

$$\pi_c^{(b)}(0) = (1 - \alpha)\frac{1}{b} + \alpha\frac{1}{b^2} = \frac{1 - \alpha + \alpha/b}{b}. \tag{6}$$

Similarly, the probability that $|C_v \setminus \{u\}| = 1$, which we denote by $\pi_c^{(b)}(1)$, is

$$\pi_c^{(b)}(1) = (1 - \alpha)\left(\frac{b-1}{b}\right) + \alpha\left[3\left(\frac{b-1}{b^2}\right)\right] = \left(\frac{b-1}{b}\right)(1 - \alpha + 3\alpha/b), \tag{7}$$

which means that either $c_v = 1$ and $u \notin C_v$, or $c_v = 2$ and either v chose u exactly once or it did not but its choices were identical. Finally, the probability that $|C_v \setminus \{u\}| = 2$, which we denote by $\pi_c^{(b)}(2)$, is the probability that $c_v = 2$ while v's choices were distinct both from each other and from u, that is,

$$\pi_c^{(b)}(2) = \alpha\left(\frac{b-1}{b}\right)\left(\frac{b-2}{b}\right). \tag{8}$$

(It is worth noting that $\pi_c^{(b)}(0)$, $\pi_c^{(b)}(1)$, and $\pi_c^{(b)}(2)$ remain as calculated even without the condition that $v \in C_u$, and in this case the use of "c" is pointless. We do insist on $v \in C_u$, however, because this is the context in which the three probabilities are used in the sequel.)

In an analogous way, let us consider a randomly chosen node u and a degree-b neighbor v of u in G such that $v \notin C_u$. If this is the case, then u and v are neighbors in D if and only if $u \in C_v$. If $c_v = 1$, then $u \in C_v$ if and only if $|C_v \setminus \{u\}| = 0$, and this happens with a probability that we denote by $\pi_{nc,1}^{(b)}(0)$ and is such that

$$\pi_{nc,1}^{(b)}(0) = 1/b. \tag{9}$$

So $1 - \pi_{nc,1}^{(b)}(0)$ is the probability that u and v are not neighbors in D, given that $v \notin C_u$ and $c_v = 1$. If $c_v = 2$, then the probability that $u \in C_v$ is $(2b - 1)/b^2$, as we see from the fact that either $|C_v \setminus \{u\}| = 0$ or $|C_v \setminus \{u\}| = 1$ may happen. The probability of the former, denoted by $\pi_{nc,2}^{(b)}(0)$, is given by

$$\pi_{nc,2}^{(b)}(0) = \frac{1}{b^2}, \tag{10}$$

while the probability of the latter, denoted by $\pi_{nc,2}^{(b)}(1)$, refers to exactly one of v's choices resulting in u. Thence

$$\pi_{nc,2}^{(b)}(1) = \frac{2b - 2}{b^2}. \tag{11}$$

So, given that $v \notin C_u$ and $c_v = 2$, $1 - \pi_{nc,2}^{(b)}(0) - \pi_{nc,2}^{(b)}(1)$ is the probability that u and v are not neighbors in D.

In the remainder of this section we analyze the efficacy of the uniform approach, concerning the performance requirements mentioned in Section 1 when D is used to disseminate I from the originator. Recall that we assume the limiting case of $n \to \infty$, so the probability that a finite-length cycle exists is negligible. We also assume that both G and D are above their phase transitions (that is, both GCC_G and GCC_D almost surely exist) and predicate our analysis upon the originator being a member of GCC_G.

2.1. Number of nodes reached

Let P_n be the ratio of the expected number of nodes reached when I is disseminated on D's edges to the expected number of nodes in GCC_G. Let also θ_G and θ_D be the fractions of n corresponding to nodes inside GCC_G and GCC_D,

respectively. There are two cases to be considered. The first case is the one in which the originator is a member of GCC_D, which occurs with probability θ_D/θ_G, this being also the ratio in this case. The second case corresponds to the originator being outside GCC_D and then the ratio is negligible. Therefore, P_n is given by

$$P_n = \frac{\theta_D^2}{\theta_G^2}. \tag{12}$$

Given a node u and one of its neighbors, say v, we define the reach of u through v in G as the set of nodes reachable by a path in G starting at u whose first edge is (u, v). We call v a dead end with respect to u in G if the reach of u through v in G is o(n). Clearly, v is a dead end with respect to u in G, and this happens with probability denoted by q, if and only if each of the other neighbors of v, in turn, is itself a dead end with respect to v in G, and this happens with probability q for each of those neighbors as well. If the degree of v is b, then it is a dead end with respect to u in G with probability q^{b-1}. And since the probability that v has degree b is given as in (1), this leads to

$$q = \sum_{b=1}^{n-1} \frac{b P_G(b)}{Z_G} q^{b-1}. \tag{13}$$

Similarly, a randomly chosen node is not in GCC_G if and only if each of its neighbors is a dead end with respect to it in G. The expected fraction of nodes inside GCC_G is then

$$\theta_G = 1 - \sum_{a=0}^{n-1} P_G(a) q^a. \tag{14}$$

The value of θ_D can be obtained in a similar, albeit more complex, way; it relies on definitions of reach and of dead-end nodes in D that are completely analogous to the ones in G. Let u be a randomly chosen node and v one of its neighbors in G. Let also q_c, $q_{nc,1}$, and $q_{nc,2}$ be the conditional probabilities that v is a dead end with respect to u in D given, respectively, that $v \in C_u$, that $v \notin C_u$ with $c_v = 1$, and that $v \notin C_u$ with $c_v = 2$. Regardless of which of these three conditions the case is, v is a dead end with respect to u in D if and only if either v is not a neighbor of u in D (which cannot happen under the first condition), or it is but the reach of u through v in D is o(n), which means that all the other neighbors of v in G are themselves dead ends with respect to it in D. We may then write

$$q_c = \sum_{b=1}^{n-1} \frac{b P_G(b)}{Z_G} \left[\pi_c^{(b)}(0)((1-\alpha)q_{nc,1} + \alpha q_{nc,2})^{b-1} + \pi_c^{(b)}(1)q_c((1-\alpha)q_{nc,1} + \alpha q_{nc,2})^{b-2} \right. $$
$$\left. + \pi_c^{(b)}(2)q_c^2((1-\alpha)q_{nc,1} + \alpha q_{nc,2})^{b-3} \right], \tag{15}$$

$$q_{nc,1} = \sum_{b=1}^{n-1} \frac{b P_G(b)}{Z_G} \left[1 - \pi_{nc,1}^{(b)}(0) + \pi_{nc,1}^{(b)}(0)((1-\alpha)q_{nc,1} + \alpha q_{nc,2})^{b-1} \right], \tag{16}$$

and

$$q_{nc,2} = \sum_{b=1}^{n-1} \frac{b P_G(b)}{Z_G} \left[1 - \pi_{nc,2}^{(b)}(0) - \pi_{nc,2}^{(b)}(1) + \pi_{nc,2}^{(b)}(0)((1-\alpha)q_{nc,1} + \alpha q_{nc,2})^{b-1} \right. $$
$$\left. + \pi_{nc,2}^{(b)}(1)q_c((1-\alpha)q_{nc,1} + \alpha q_{nc,2})^{b-2} \right], \tag{17}$$

referring back to the probabilities calculated in the introduction to Section 2.

Each of the expressions in (15)–(17) illustrates our use of those probabilities. We comment on (15) in detail and urge the reader to consider each of the others, as well as other expressions yet to come, in a similar light. In (15), and for $k \in \{0, 1, 2\}$, $q_c^k((1-\alpha)q_{nc,1} + \alpha q_{nc,2})^{b-1-k}$ gives the conditional probability that all the $b-1$ neighbors of a degree-b neighbor v of u that are not u are dead ends with respect to v in D; the condition is that $|C_v \setminus \{u\}| = k$, which happens with probability $\pi_c^{(b)}(k)$. Thus q_c^k is the dead-end probability for the k neighbors in $C_v \setminus \{u\}$, and $((1-\alpha)q_{nc,1} + \alpha q_{nc,2})$ is the dead-end probability for each of the $b - 1 - k$ neighbors that are not in $C_v \cup \{u\}$.

Now, since a randomly chosen node is not in GCC_D if and only if each of its neighbors in G is a dead end with respect to it in D, we have

$$\theta_D = 1 - P_G(0) - \sum_{a=1}^{n-1} P_G(a) \left[\pi_r^{(a)}(1) q_c((1-\alpha)q_{\text{nc},1} + \alpha q_{\text{nc},2})^{a-1} + \pi_r^{(a)}(2) q_c^2((1-\alpha)q_{\text{nc},1} + \alpha q_{\text{nc},2})^{a-2} \right]. \tag{18}$$

Notice that, in (18), $P_G(0)$ must be singled out of the sum in order to be taken into account, since $\pi_r^{(0)}(1) = \pi_r^{(0)}(2) = 0$.

2.2. Number of messages sent

We denote by P_m the ratio of the expected number of messages sent when disseminating I through D's edges to the expected number of messages sent when disseminating I through the edges of a spanning tree of GCC_G. Let Z_{GCC_D} be the average degree of nodes in D conditioned upon membership in GCC_D. Restricting our analysis to the case in which the originator is in GCC_D (the other case leads to a negligible ratio), which occurs with probability θ_D/θ_G, the expected number of messages sent on the edges of D is $n\theta_D Z_{\text{GCC}_D}$, while the expected number of messages sent on the edges of the spanning tree of GCC_G is $2(n\theta_G - 1)$, thus leading to

$$P_m = \frac{\theta_D^2 n Z_{\text{GCC}_D}}{2\theta_G(n\theta_G - 1)}. \tag{19}$$

Now let $P_D(i \mid \text{GCC}_D)$ be the probability that a randomly chosen node of GCC_D has degree i in D. We have

$$Z_{\text{GCC}_D} = \sum_{i=0}^{n-1} i P_D(i \mid \text{GCC}_D). \tag{20}$$

If also we let $P_D(i \mid a, \text{GCC}_D)$ be the conditional probability that a randomly chosen node of GCC_D has degree i in D given that it has degree a in G, then $P_D(i \mid \text{GCC}_D)$ can be written as

$$P_D(i \mid \text{GCC}_D) = \sum_{a=i}^{n-1} P_D(i \mid a, \text{GCC}_D) P_G(a \mid \text{GCC}_D), \tag{21}$$

where $P_G(a \mid \text{GCC}_D)$ is the probability that the node has degree a in G. We henceforth approximate $P_D(i \mid a, \text{GCC}_D)$ by $P_D(i \mid a)$, which is the probability that a randomly chosen, degree-a node of G has degree i in D (more on this approximation in Section 2.3). Also, if u is this node and v one of its neighbors in G such that $v \notin C_u$, then the probability that $u \in C_v$, denoted by r, is given by

$$r = \sum_{b=1}^{n-1} \frac{b P_G(b)}{Z_G} \left[(1-\alpha)\pi_{\text{nc},1}^{(b)}(0) + \alpha(\pi_{\text{nc},2}^{(b)}(0) + \pi_{\text{nc},2}^{(b)}(1)) \right] \tag{22}$$

and yields

$$P_D(i \mid a) = \pi_r^{(a)}(1)\binom{a-1}{i-1}r^{i-1}(1-r)^{a-i} + \pi_r^{(a)}(2)\binom{a-2}{i-2}r^{i-2}(1-r)^{a-i}. \tag{23}$$

Using Bayes' rule to rewrite $P_G(a \mid \text{GCC}_D)$ as

$$P_G(a \mid \text{GCC}_D) = \frac{P_D(\text{GCC}_D \mid a) P_G(a)}{P_D(\text{GCC}_D)}, \tag{24}$$

where $P_D(\text{GCC}_D) = \theta_D$ and

$$P_D(\text{GCC}_D \mid a) = 1 - \pi_r^{(a)}(1) q_c((1-\alpha)q_{\text{nc},1} + \alpha q_{\text{nc},2})^{a-1} - \pi_r^{(a)}(2) q_c^2((1-\alpha)q_{\text{nc},1} + \alpha q_{\text{nc},2})^{a-2}, \tag{25}$$

leads, finally, to

$$Z_{\mathrm{GCC}_D} = \sum_{i=0}^{n-1} i \sum_{a=i}^{n-1} P_D(i \mid a) \frac{P_D(\mathrm{GCC}_D \mid a) P_G(a)}{\theta_D}$$

$$= \sum_{a=0}^{n-1} \frac{P_D(\mathrm{GCC}_D \mid a) P_G(a)}{\theta_D} \sum_{i=0}^{a} i P_D(i \mid a), \tag{26}$$

where

$$\sum_{i=0}^{a} i P_D(i \mid a) = \pi_{\mathrm{r}}^{(a)}(1)(1 + (a-1)r) + \pi_{\mathrm{r}}^{(a)}(2)(2 + (a-2)r). \tag{27}$$

2.3. Node degree

Let us now consider the expected degree of D conditioned upon membership in GCC_G, which we denote by Z_{D,GCC_G}. Let $P_D(i \mid \mathrm{GCC}_G)$ be the probability that a randomly chosen node of GCC_G has degree i in D. Then Z_{D,GCC_G} is clearly given by

$$Z_{D,\mathrm{GCC}_G} = \sum_{i=0}^{n-1} i P_D(i \mid \mathrm{GCC}_G), \tag{28}$$

where

$$P_D(i \mid \mathrm{GCC}_G) = \sum_{a=i}^{n-1} P_D(i \mid a, \mathrm{GCC}_G) P_G(a \mid \mathrm{GCC}_G), \tag{29}$$

$P_D(i \mid a, \mathrm{GCC}_G)$ being the probability that a randomly chosen, degree-a node of GCC_G has degree i in D. This lets (28) be rewritten as

$$Z_{D,\mathrm{GCC}_G} = \sum_{i=0}^{n-1} i \sum_{a=i}^{n-1} P_D(i \mid a, \mathrm{GCC}_G) P_G(a \mid \mathrm{GCC}_G)$$

$$= \sum_{a=0}^{n-1} P_G(a \mid \mathrm{GCC}_G) \sum_{i=0}^{a} i P_D(i \mid a, \mathrm{GCC}_G), \tag{30}$$

which, using Bayes' rule to write

$$P_G(a \mid \mathrm{GCC}_G) = \frac{P_G(\mathrm{GCC}_G \mid a) P_G(a)}{P_G(\mathrm{GCC}_G)} = \frac{(1 - q^a) P_G(a)}{\theta_G} \tag{31}$$

(cf. (13)), yields

$$Z_{D,\mathrm{GCC}_G} = \sum_{a=0}^{n-1} \frac{(1 - q^a) P_G(a)}{\theta_G} \sum_{i=0}^{a} i P_D(i \mid a, \mathrm{GCC}_G). \tag{32}$$

One possibility now would be to proceed similarly to what we did in Section 2.2 and approximate $P_D(i \mid a, \mathrm{GCC}_G)$ by $P_D(i \mid a)$. This would immediately let us use (27) in (32) and be done. However, we know from early experiments like the ones to be discussed in Section 3 that, unlike the case of Section 2.2, this sometimes yields an agreement between analytical prediction and simulation that is not satisfactory. In the present case, then, we look more closely at the nature of $P_D(i \mid a, \mathrm{GCC}_G)$ and first notice that (23)—and consequently (27) as well—can essentially be used to express $P_D(i \mid a, \mathrm{GCC}_G)$, provided the probability r appearing in it is made to depend on a and on the degree-a node's membership in GCC_G. This is to be taken in opposition to the expression for r in (22), but clearly all that needs to be changed in (22) to make the dependencies manifest is to replace $b P_G(b)/Z_G$ as the probability that a given neighbor of a degree-a node has degree b. The reason why this is so is that, given a degree-a node's membership in GCC_G, that probability is no longer independent from a (as we know from our comment following (3), the existence of GCC_G is related to nodes' neighbors' degrees).

Let p be the probability that we seek for use in place of $b P_G(b)/Z_G$. That is, p is the probability that a given neighbor of a degree-a node of GCC_G has degree b. The probability that a degree-a node is in GCC_G is $1 - q^a$ (all its a neighbors

must otherwise be dead ends with respect to it in G), so the probability that a degree-a node is a member of GCC_G and moreover a given neighbor of it has degree b is $p(1 - q^a)$. But this latter probability can also be expressed as $(bP_G(b)/Z_G)(1 - q^{a+b-2})$, since the rightmost factor is the probability of the condition that a pair of neighbors (one of degree a, the other of degree b) is in GCC_G, and furthermore $bP_G(b)/Z_G$, under that condition, continues to give the probability that a given neighbor of a degree-a node has degree b. We may then write

$$p(1 - q^a) = \frac{bP_G(b)}{Z_G}(1 - q^{a+b-2}), \tag{33}$$

from which it follows that

$$p = \frac{bP_G(b)}{Z_G}\left(\frac{1 - q^{a+b-2}}{1 - q^a}\right). \tag{34}$$

We note, finally, that an analogous but more complicated development could also be used to avoid the approximation of $P_D(i \mid a, \mathrm{GCC}_D)$ by $P_D(i \mid a)$ made in Section 2.2. As explained, however, that would only add needless detail.

2.4. Path length

Let P_t be the ratio of the expected path length from the originator in D to the expected path length from the originator in G. We may again restrict our analysis to the case in which the originator is a member of GCC_D, which occurs with probability θ_D/θ_G. Denoting by L_{GCC_G} and L_{GCC_D} the average path lengths from the originator in G and in D, respectively, we have

$$P_t = \frac{\theta_D L_{\mathrm{GCC}_D}}{\theta_G L_{\mathrm{GCC}_G}}. \tag{35}$$

By (31), the average degree in G of the nodes inside GCC_G, denoted by Z_{GCC_G}, is

$$Z_{\mathrm{GCC}_G} = \sum_{a=0}^{n-1} \frac{a(1 - q^a)P_G(a)}{\theta_G}. \tag{36}$$

Now let two nodes be called ℓ-neighbors in a graph, for $\ell \geqslant 1$, when the distance between them in the graph is ℓ (the case of $\ell = 1$ is simply the case of neighbors in the graph). Given a randomly chosen node u in GCC_G and one of its neighbors, say v, the expected number of v's other neighbors (i.e., excluding u), denoted by ρ, is [1]

$$\rho = \sum_{b=1}^{n-1} \frac{bP_G(b)}{Z_G}(b - 1), \tag{37}$$

and then the expected number of u's 2-neighbors in G, which we denote by $Z^{(2)}_{\mathrm{GCC}_G}$, is

$$Z^{(2)}_{\mathrm{GCC}_G} = Z_{\mathrm{GCC}_G}\rho. \tag{38}$$

In general, the expected number of u's ℓ-neighbors in G, denoted by $Z^{(\ell)}_{\mathrm{GCC}_G}$, is

$$Z^{(\ell)}_{\mathrm{GCC}_G} = Z^{(\ell-1)}_{\mathrm{GCC}_G}\rho = Z_{\mathrm{GCC}_G}\rho^{\ell-1}. \tag{39}$$

We can then obtain an approximation for L_{GCC_G} by summing the values of $Z^{(\ell)}_{\mathrm{GCC}_G}$ from $\ell = 1$ up until the sum becomes equal to the expected number of nodes inside GCC_G minus one (to account for node u). Thus

$$\sum_{\ell=1}^{L_{\mathrm{GCC}_G}} Z^{(\ell)}_{\mathrm{GCC}_G} = n\theta_G - 1, \tag{40}$$

---

[1] The use of $bP_G(b)/Z_G$ in (37) and later in (42)–(46) is in principle subject to the same corrections explained at the end of Section 2.3. However, introducing those corrections in the present context not only seems unnecessary given the computational results to be discussed in Section 3, but also would lead to much more complicated (and probably insoluble) versions of (40) and (56).

which yields

$$L_{\text{GCC}_G} = \frac{\ln[((n\theta_G - 1)/Z_{\text{GCC}_G})(\rho - 1) + 1]}{\ln \rho}. \tag{41}$$

Let us now turn to obtaining the value of L_{GCC_D}. Consider a randomly chosen node u of GCC_D and a neighbor v of u in G. If $v \notin C_u$, then let $r_{\text{nc},1}$ be the conditional probability that $u \in C_v$ given that $c_v = 1$. We have

$$r_{\text{nc},1} = \sum_{b=1}^{n-1} \frac{b P_G(b)}{Z_G} \pi_{\text{nc},1}^{(b)}(0). \tag{42}$$

Likewise, if $r_{\text{nc},2}$ is the conditional probability that $u \in C_v$ given that $c_v = 2$, then

$$r_{\text{nc},2} = \sum_{b=1}^{n-1} \frac{b P_G(b)}{Z_G} \left(\pi_{\text{nc},2}^{(b)}(0) + \pi_{\text{nc},2}^{(b)}(1) \right). \tag{43}$$

Calculating the expected number of other neighbors of a neighbor v of u requires three cases to be considered. The first case is the case of $v \in C_u$, and then the expected number of u's 2-neighbors in D that are reachable from v is $t_c(1, r_{\text{nc},1}, r_{\text{nc},2})$, where

$$t_c(x, y, z) = \sum_{b=1}^{n-1} \frac{b P_G(b)}{Z_G} \left\{ \pi_c^{(b)}(0)(b-1)((1-\alpha)y + \alpha z) + \pi_c^{(b)}(1)[x + (b-2)((1-\alpha)y + \alpha z)] \right.$$
$$\left. + \pi_c^{(b)}(2)[2x + (b-3)((1-\alpha)y + \alpha z)] \right\}. \tag{44}$$

In the second case, $v \notin C_u$ with $c_v = 1$. We similarly let

$$t_{\text{nc},1}(y, z) = \sum_{b=1}^{n-1} \frac{b P_G(b)}{Z_G} \left[\pi_{\text{nc},1}^{(b)}(0)(b-1)((1-\alpha)y + \alpha z) \right], \tag{45}$$

and then the expected number of u's 2-neighbors in D that are reachable from v is $t_{\text{nc},1}(r_{\text{nc},1}, r_{\text{nc},2})$.

The third and final case is that of $v \notin C_u$ with $c_v = 2$. As in the previous two cases, we let

$$t_{\text{nc},2}(x, y, z) = \sum_{b=1}^{n-1} \frac{b P_G(b)}{Z_G} \left\{ \pi_{\text{nc},2}^{(b)}(0)(b-1)((1-\alpha)y + \alpha z) + \pi_{\text{nc},2}^{(b)}(1)[x + (b-2)((1-\alpha)y + \alpha z)] \right\}, \tag{46}$$

which yields the expected number of u's 2-neighbors in D that are reachable from v as $t_{\text{nc},2}(1, r_{\text{nc},1}, r_{\text{nc},2})$.

Now, for u a randomly chosen node in GCC_D, and recalling (24), letting

$$t_r(x, y, z) = \sum_{a=0}^{n-1} \frac{P_D(\text{GCC}_D \mid a) P_G(a)}{\theta_D} \left\{ \pi_r^{(a)}(1)[x + (a-1)((1-\alpha)y + \alpha z)] \right.$$
$$\left. + \pi_r^{(a)}(2)[2x + (a-2)((1-\alpha)y + \alpha z)] \right\} \tag{47}$$

allows the expected degree of u to be expressed as $t_r(1, r_{\text{nc},1}, r_{\text{nc},2})$, and similarly the expected number of 2-neighbors of u as

$$t_r(t_c(1, r_{\text{nc},1}, r_{\text{nc},2}), t_{\text{nc},1}(r_{\text{nc},1}, r_{\text{nc},2}), t_{\text{nc},2}(1, r_{\text{nc},1}, r_{\text{nc},2})). \tag{48}$$

For simplicity's sake, let β_c^c, $\beta_c^{\text{nc},1}$, $\beta_c^{\text{nc},2}$, $\beta_{\text{nc},1}^{\text{nc},1}$, $\beta_{\text{nc},1}^{\text{nc},2}$, $\beta_{\text{nc},2}^c$, $\beta_{\text{nc},2}^{\text{nc},1}$, $\beta_{\text{nc},2}^{\text{nc},2}$, β_r^c, $\beta_r^{\text{nc},1}$, and $\beta_r^{\text{nc},2}$ be such that (44), (45), (46), and (47) can, respectively, be rewritten as

$$t_c(x, y, z) = \beta_c^c x + \beta_c^{\text{nc},1} y + \beta_c^{\text{nc},2} z, \tag{49}$$

$$t_{\mathrm{nc},1}(y, z) = \beta_{\mathrm{nc},1}^{\mathrm{nc},1} y + \beta_{\mathrm{nc},1}^{\mathrm{nc},2} z, \tag{50}$$

$$t_{\mathrm{nc},2}(x, y, z) = \beta_{\mathrm{nc},2}^{\mathrm{c}} x + \beta_{\mathrm{nc},2}^{\mathrm{nc},1} y + \beta_{\mathrm{nc},2}^{\mathrm{nc},2} z, \tag{51}$$

and

$$t_{\mathrm{r}}(x, y, z) = \beta_{\mathrm{r}}^{\mathrm{c}} x + \beta_{\mathrm{r}}^{\mathrm{nc},1} y + \beta_{\mathrm{r}}^{\mathrm{nc},2} z. \tag{52}$$

Then, introducing the row vector $A = \begin{bmatrix} \beta_{\mathrm{r}}^{\mathrm{c}} & \beta_{\mathrm{r}}^{\mathrm{nc},1} & \beta_{\mathrm{r}}^{\mathrm{nc},2} \end{bmatrix}$ and the matrix

$$\mathrm{B} = \begin{bmatrix} \beta_{\mathrm{c}}^{\mathrm{c}} & \beta_{\mathrm{c}}^{\mathrm{nc},1} & \beta_{\mathrm{c}}^{\mathrm{nc},2} \\ 0 & \beta_{\mathrm{nc},1}^{\mathrm{nc},1} & \beta_{\mathrm{nc},1}^{\mathrm{nc},2} \\ \beta_{\mathrm{nc},2}^{\mathrm{c}} & \beta_{\mathrm{nc},2}^{\mathrm{nc},1} & \beta_{\mathrm{nc},2}^{\mathrm{nc},2} \end{bmatrix},$$

we have that the expected number of ℓ-neighbors of a randomly chosen node in GCC_D, which for $\ell \geqslant 1$ we denote by $Z_{\mathrm{GCC}_D}^{(\ell)}$, is

$$Z_{\mathrm{GCC}_D}^{(\ell)} = A\mathrm{B}^{\ell-1} \begin{bmatrix} 1 \\ r_{\mathrm{nc},1} \\ r_{\mathrm{nc},2} \end{bmatrix}. \tag{53}$$

(The reader should check that (53) yields the Z_{GCC_D} of (26) for $\ell = 1$, and also that it becomes (48) for $\ell = 2$.)

We are now left with the task of finally obtaining L_{GCC_D}. This can only be achieved numerically, and to this end we resort to the eigenvalues (say λ_1, λ_2, and λ_3, which we assume are all distinct [2]) and corresponding eigenvectors (v_1, v_2, and v_3) of B. If V is the matrix whose columns are v_1, v_2, and v_3, and Λ the matrix having diagonal elements λ_1, λ_2, and λ_3 with 0's everywhere else, then B can be diagonalized into Λ via

$$\Lambda = V^{-1}\mathrm{B}V, \tag{54}$$

which can be equivalently expressed as

$$\mathrm{B} = V\Lambda V^{-1} \tag{55}$$

and used to obtain the $\mathrm{B}^{\ell-1}$ of (53) as $\mathrm{B}^{\ell-1} = V\Lambda^{\ell-1}V^{-1}$ [12].

Proceeding in a manner analogous to the one that led to (40), we can then obtain L_{GCC_D} by numerically solving the equation

$$\sum_{\ell=1}^{L_{\mathrm{GCC}_D}} Z_{\mathrm{GCC}_D}^{(\ell)} = n\theta_D - 1, \tag{56}$$

which, by (53), is equivalent to

$$AV \begin{bmatrix} \dfrac{\lambda_1^{L_{\mathrm{GCC}_D}} - 1}{\lambda_1 - 1} & 0 & 0 \\ 0 & \dfrac{\lambda_2^{L_{\mathrm{GCC}_D}} - 1}{\lambda_2 - 1} & 0 \\ 0 & 0 & \dfrac{\lambda_3^{L_{\mathrm{GCC}_D}} - 1}{\lambda_3 - 1} \end{bmatrix} V^{-1} \begin{bmatrix} 1 \\ r_{\mathrm{nc},1} \\ r_{\mathrm{nc},2} \end{bmatrix} = n\theta_D - 1. \tag{57}$$

---

[2] This has proven true in all the experimental scenarios of Section 3, so we dwell on the matter no further.

3. Simulation results

In addition to our mathematical analysis of Section 2, and seeking to validate it experimentally, we have carried out simulations of the uniform approach to the construction of D. Also, and notwithstanding the fact that our analysis has not included the degree-based approach, we have extended our simulations to cover it as well. Each of our simulations is based on disseminating I on random graphs, which are always generated to be above the corresponding phase transition. This allows the simulation to be constrained to operate within the graph's largest connected component, which almost surely is a giant connected component.

We have considered two random-graph models. The first model corresponds to the classical model of Erdős and Rényi [9], in which G is constructed on n nodes by letting each of the possible $n(n - 1)/2$ edges exist with constant probability $z/(n - 1)$ for $0 < z \leqslant n - 1$. As a result, G has node degrees distributed according to a Poisson distribution, that is, the probability that a node has degree a is $P_G(a) = \mathrm{e}^{-z}z^a/a!$ [5]. For Poisson-distributed node degrees, it follows from (2) that the graph is above the phase transition if and only if $z > 1$, since $Z_G = z$ and $\langle K_G^2 \rangle = z^2 + z$.

We have concentrated on analyzing the behavior of P_n, P_m, Z_{D,GCC_G}, and P_t for $1 \leqslant z \leqslant 10$ and $\alpha = 0.10, 0.25, 0.50,$ 0.75, 1.00. To this end, and for each value of z, we generated 300 random graphs with $n = 10\,000$ nodes, and then constructed two instances of D for each value of α, one following the uniform approach and the other the degree-based approach. On each of the instances, we then conducted 1000 disseminations by randomly choosing an originator from among the nodes in the largest connected component of G. At the end, we averaged the quantities of interest overall to obtain P_n, P_m, Z_{D,GCC_G}, and P_t. (Note that obtaining Z_{D,GCC_G} does not depend on any dissemination, but rather only on the available G and D instances. The same is in principle also true of P_t, but we simulate disseminations by breadth-first search from the originator and P_t is then approximated along the way.)

Fig. 1 shows simulation results for random graphs having Poisson-distributed node degrees. Parts (a–d), concerning the uniform approach, show an excellent agreement between analytical and simulation results, with only a slight deviation in part (d), which is in all likelihood to be attributed to the approximations made in Section 2.4. The plots for P_n (Fig. 1(a, e)) evidence the expected superiority of the degree-based approach over the uniform approach, since in the former case P_n approaches 1 rapidly as z is increased, more or less regardless of α (in the uniform approach, this only seems to happen for $z < 10$ when $\alpha \geqslant 0.50$). A closer examination of the data for the degree-based approach, say for $z = 5$ and $\alpha = 0.50$, reveals $P_n \approx 0.998$, $P_m \approx 1.24$, $Z_{D,\mathrm{GCC}_G} \approx 2.48$, and $P_t \approx 1.76$. What this means is that, using roughly 1.24 times as many edges as a spanning tree and paths that, on average, are greater than those of G by a factor of only 1.76, the dissemination subgraph reaches almost all the nodes of the network while having a relatively low average node degree. Comparing the two approaches, it is curious to note that the plots of P_m (Fig. 1(b, f)) are very similar to each other, the same holding for those of Z_{D,GCC_G} (Fig. 1(c, g)), which indicates that the number of edges in the dissemination subgraph is quite independent of whether one approach is used or the other. However, the difference between the P_n plots demonstrates that the choices made by the nodes in the degree-based approach somehow lead the edges to end up deployed in such a manner as to favor the connectedness of the dissemination subgraph strongly.

The other random-graph model we have considered is the one in which node degrees are distributed according to a power law. The probability that a node in G has degree a is in this case, and for $n \to \infty$, given by $P_G(a) = a^{-\tau}/\zeta(\tau)$, where $\tau > 1$ is a parameter and $\zeta(x)$ is the Riemann zeta function [18], that is, $\zeta(x) = \sum_{y=1}^{\infty} y^{-x}$. Then we have $Z_G = \zeta(\tau - 1)/\zeta(\tau)$ and $\langle K_G^2 \rangle = \zeta(\tau - 2)/\zeta(\tau)$, so solving (2) numerically yields $\tau < 3.47$ as the condition for G to be above the phase transition. We have performed simulations for $2 \leqslant \tau \leqslant 3$ in the same way as we did for the Poisson case.

Random graphs with degrees thus distributed can be generated in two phases. First the degrees $a_1, a_2, \ldots, a_n$ of the n nodes, constituting the graph's so-called degree sequence, are sampled repeatedly from the power law until $\sum_{i=1}^{n} a_i$ comes out even. Then $\sum_{i=1}^{n} a_i$ labeled balls are put inside an imaginary urn, where exactly a_i of the balls are labeled i, for $1 \leqslant i \leqslant n$. A pair of balls, say of labels u and v, is then withdrawn from the urn and the edge (u, v) is added to the graph; this process is repeated until the urn becomes empty. This algorithm clearly generates a multigraph, where self-loops and multiple edges are allowed to exist. What we do as a last step is to discard such undesirable edges, which at the end yields a random graph whose degree sequence is an approximation of the one sampled.
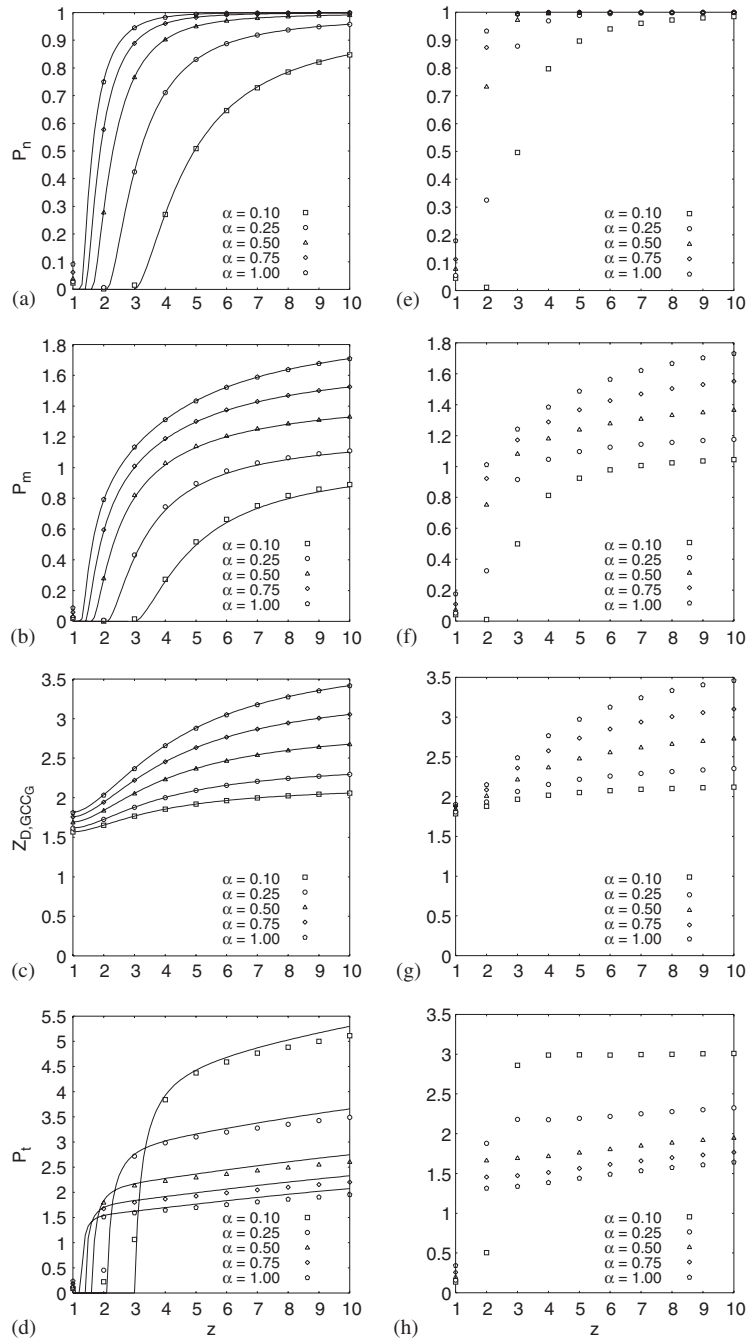
Fig. 1. Simulation results for the uniform (a–d) and the degree-based (e–h) approach on random graphs having Poisson-distributed node degrees. The plots show P_n (a, e), P_m (b, f), Z_{D,GCC_G} (c, g), and P_t (d, h) for $\alpha = 0.10, 0.25, 0.50, 0.75, 1.00$. Solid lines give the analytical predictions of Section 2.

Fig. 2 shows simulation results for random graphs having node degrees distributed according to a power law. The plots for the uniform approach (Fig. 2(a–d)) show poor results, as P_n stays clear of 1 for most values of τ, but they do nonetheless corroborate the analytical predictions of Section 2 in parts (a), (b), and (c). For part (d) no analytical result is given, since Eqs. (39) and (53), as similarly observed in [16], do not converge. The plots for the degree-
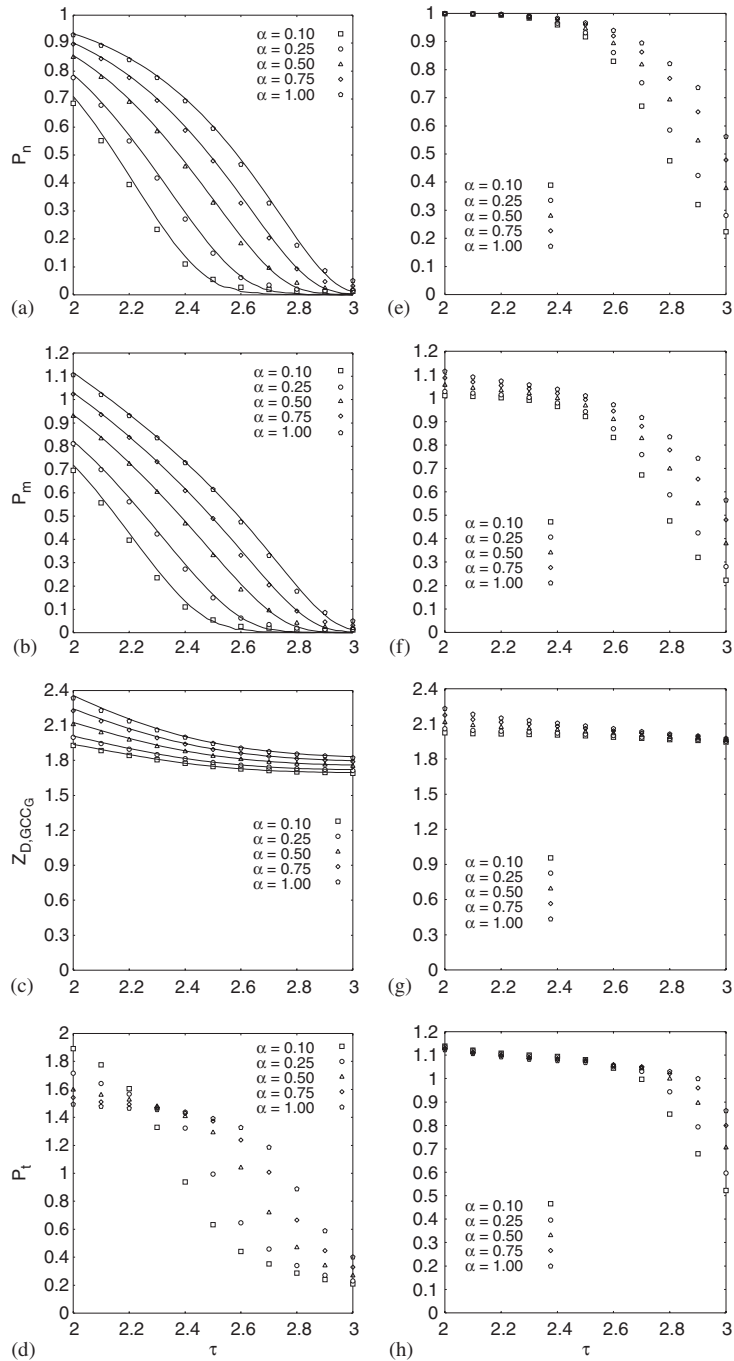
Fig. 2. Simulation results for the uniform (a–d) and the degree-based (e–h) approach on random graphs having node degrees distributed as a power law. The plots show P_n (a, e), P_m (b, f), Z_{D,GCC_G} (c, g), and P_t (d, h) for $\alpha = 0.10, 0.25, 0.50, 0.75, 1.00$. Solid lines give the analytical predictions of Section 2.

based approach (Fig. 2(e–h)), in turn, show excellent results for $\tau < 2.4$. In this range, $P_n \approx 1$ and both P_m and P_t are slightly above 1, regardless of the value of α, thus demonstrating that the dissemination subgraph is very close to a spanning tree. As for Z_{D,GCC_G}, it stays modestly valued below roughly 2.25 throughout the entire spectrum of τ values.

4. Resilience and adaptability

4.1. Resilience to node and link failures

Let γ be the probability that a given transmission by a node over a link is successful. This probability is the product of two others, namely that the node is operational and that the link is operational. We now consider the problem of using a dissemination subgraph to disseminate information when each transmission has a failure probability of $1 - \gamma$. (Note, before we begin, that a simple protocol employing acknowledgement messages to ensure reliable transmissions can be used when γ is substantially low. In spite of this fact, our interest is to verify what happens to the value of P_n when a failure may occur and no additional message is sent to make up for it.)

Let us consider what happens to GCC_G when failures may occur. For such, let G' be the graph obtained from G by independently removing every edge with probability $1 - \gamma$. Employing the same nomenclature as in Section 2, a node of G' is outside $\mathrm{GCC}_{G'}$ if and only if each of its neighbors in G is a dead end with respect to it in G'. Considering a randomly chosen node u, let q' be the probability that a given neighbor v of u in G is a dead end with respect to it in G'. We have

$$q' = \sum_{b=1}^{n-1} \frac{b P_G(b)}{Z_G} (1 - \gamma + \gamma q')^{b-1}, \tag{58}$$

where $(1 - \gamma + \gamma q')^{b-1}$ indicates, when b is the degree of v, that each of the neighbors of v in G that are not u either is not a neighbor of v in G' or is itself a dead end with respect to v in G'. So, if $\theta_{G'}$ is the expected size of $\mathrm{GCC}_{G'}$, we obtain

$$\theta_{G'} = 1 - \sum_{a=0}^{n-1} P_G(a)(q')^a. \tag{59}$$

We have carried out simulations on random graphs having $n = 10\,000$ and degrees distributed according to either a Poisson distribution with $1 \leqslant z \leqslant 10$ or a power law with $2 \leqslant \tau \leqslant 3$. Our aim has been to analyze P_n in the degree-based approach when $\gamma = 0.95$, i.e., when each transmission has a 0.95 probability of success. These simulations have followed the same methodology as in Section 3.

Notice that an upper bound on P_n when $\gamma > 0$ can be obtained by considering a dissemination on all the edges of G'. Such a bound is thus $\theta_{G'}^2/\theta_G^2$. The degree-based approach to the construction of D will then be as resilient to failures as P_n is close to $\theta_{G'}^2/\theta_G^2$.

Fig. 3 shows the results for $\alpha = 0.50, 0.75, 1.00$ and provides an indication of how resilient the dissemination subgraph is to transmission failures. Clearly, in both the Poisson case (part (a) of the figure) and the power-law case (part (b)), P_n approaches $\theta_{G'}^2/\theta_G^2$ as G gets denser (i.e., higher z or lower τ, as the case may be).
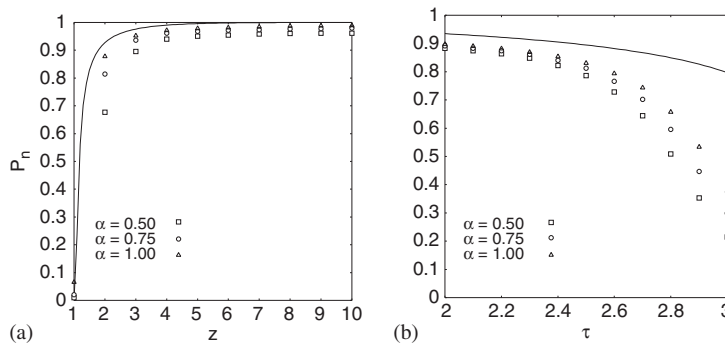

Fig. 3. Simulation results of the degree-based approach with $\gamma = 0.95$ on random graphs having Poisson (a) and power-law (b) node-degree distributions for $\alpha = 0.50, 0.75, 1.00$. Solid lines give the analytical predictions for $\theta_{G'}^2/\theta_G^2$.

4.2. Adaptability to topology changes

We now take a brief look at how a dissemination subgraph D can be made to cope with dynamic topology changes in G. As customary in such cases, we model the addition or removal of a node as, respectively, the addition or removal of the edges that are incident to it. It then suffices that we consider the addition or removal of single edges, in which context we further assume that the two end nodes of the edge in question are capable of detecting its appearance or disappearance instantaneously.

The crux of this adaptability issue is that D, being constructed by strictly local actions by the nodes, can undergo changes that affect only a small vicinity of the edge that is being added or removed (this is to be contrasted with other situations—cf., e.g., [1]—in which the impact of topological changes spreads much more widely). Let (u, v) be an edge that is added to or removed from G. In the uniform approach, only u and v need remake their choices; in the degree-based approach, this holds for u and v, and also for their neighbors (whose choices are affected by the degree of u or v, as the case may be).

5. Conclusions

In this paper we have considered the use of a spanning subgraph for disseminating a piece of information, originally known to a single node, to all the other nodes of an unstructured network. We have introduced two local heuristics, referred to as the uniform and the degree-based approach, for building what we call a dissemination subgraph. As we argued toward the end of the paper, the heuristics' intrinsically local nature leads to a degree of resilience of the dissemination subgraph to failures, and also to a relative ease of adaptation to topological changes.

We have contributed an innovative mathematical analysis of the uniform approach, one that we hope can be extended to the degree-based approach as well, and also inspire the mathematical analysis of similar problems. Our simulations on random graphs corroborate our analytical results for the uniform approach and demonstrate the efficacy, in terms of some relevant indicators, of the degree-based approach for networks in which node degrees are distributed according to a Poisson distribution or to a power law.

We find it remarkable that independent, strictly local decisions by the nodes of a complex network are capable of giving rise to a global structure that in many cases comes very near a subgraph with, on average, important properties related to its use as a substrate for information dissemination. These properties include the ability to reach nearly every node in the originator's connected component in the network, and do so with relatively modest requirements concerning the overall number of messages and per-node transmission bandwidth. They also include stretching paths only by a small factor when compared to the corresponding paths in the network.

References

[1] Y. Afek, D. Hendler, On the complexity of global computation in the presence of link failures: the general case, Distributed Comput. 8 (1995) 115–120.
[2] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (2002) 47–97.
[3] B. Awerbuch, Optimal distributed algorithms for minimum weight spanning tree, counting, leader election, and related problems, in: Proc. 19th Annu. ACM Conf. Theory of Computing, 1987, pp. 230–240.
[4] V.C. Barbosa, An Introduction to Distributed Algorithms, The MIT Press, Cambridge, MA, 1996.
[5] B. Bollobás, Random Graphs, second ed., Cambridge University Press, Cambridge, UK, 2001.
[6] C. Cheng, I.A. Cimet, S.P.R. Kumar, A protocol to maintain a minimum spanning tree in a dynamic topology, in: Sympos. Proc. on Communications Architectures and Protocols, 1988, pp. 330–337.
[7] Y.-H. Chu, S.G. Rao, H. Zhang, A case for end system multicast, in: Proc. 2000 ACM SIGMETRICS Internat. Conf. Measurement and Modeling of Computer Systems, 2000, pp. 1–12.
[8] R. Cohen, K. Erez, D. ben Avraham, S. Havlin, Resilience of the Internet to random breakdowns, Phys. Rev. Lett. 85 (2000) 4626.
[9] P. Erdős, A. Rényi, On random graphs, Publ. Math. 6 (1959) 290–297.

[10] M. Faloutsos, P. Faloutsos, C. Faloutsos, On power-law relationships of the Internet topology, in: Proc. Conf. Applications, Technologies, Architectures, and Protocols for Computer Communication, 1999, pp. 251–262.

[11] M. Faloutsos, M. Molle, Optimal distributed algorithm for minimum spanning trees revisited, in: Proc. 14th Annu. ACM Sympos. Principles of Distributed Computing, 1995, pp. 231–237.

[12] S.H. Friedberg, A.J. Insel, L.E. Spence, Linear Algebra, fourth ed., Prentice-Hall, Upper Saddle River, NJ, 2003.

[13] G. Kortsarz, D. Peleg, Generating low-degree 2-spanners, SIAM J. Comput. 30 (1998) 1438–1456.

[14] A. Medina, I. Matta, J. Byers, On the origin of power laws in Internet topologies, Comput. Commun. Rev. 30 (2) (2000) 18–28.

[15] M. Molloy, B. Reed, A critical point for random graphs with a given degree sequence, Random Structures and Algorithms 6 (1995) 161–180.

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

[17] D. Peleg, Distributed Computing, SIAM, Philadelphia, PA, 2000.

[18] S.Y. Yan, Number Theory for Computing, second ed., Springer, Berlin, Germany, 2002.