Robustness in Nonorthogonal Multiple Access 5G Networks

Benjamin Pimentel
Naval Postgraduate School
benjamin.pimentel@nps.edu

Alex Bordetsky
Naval Postgraduate School
abordets@nps.edu

Ralucca Gera
Naval Postgraduate School
rgera@nps.edu

Abstract

The diversity of fifth generation (5G) network use cases, multiple access technologies, and network deployments requires measures of network robustness that complement throughput-centric error rates. In this paper, we investigate robustness in nonorthogonal multiple access (NOMA) 5G networks through temporal network theory. We develop a graph model and analytical framework to characterize time-varying network connectedness as a function of NOMA overloading. We extend our analysis to derive lower bounds and probability expressions for the number of medium access control frames required to achieve pairwise connectivity between all network devices. We support our analytical results through simulation.

1. Introduction

The pursuit of 5G wireless network performance objectives in [1] has led to many advances in radio access network technology including nonorthogonal multiple access (NOMA) [2, 3]. NOMA is a family of multiple access approaches that assign multiple network devices (ND) to transmit or receive radio frequency (RF) signals using nonorthogonal time-frequency resources, or resource blocks (RB). Unlike orthogonal multiple access (OMA) schemes that separate ND transmissions in frequency, time, code, space, or some combination thereof, NOMA allows simultaneous use of RBs, and employs a variety of multiuser detection (MUD) methods to separate the signals [4]. NOMA methods are broadly categorized as power-domain or code-domain. Power-domain approaches differentiate between NDs using a combination of superposition coding and successive interference cancellation while code-domain schemes achieve MUD through sparse or nonorthogonal low cross-correlation spreading sequences [5, 6, 7, 8].

A ratio of NDs to RBs that is greater than one is called *overloading*. Through overloading, NOMA has the potential to support increased aggregate throughput,

improved spectral efficiency, lower latency, and higher connection densities than OMA [6, 7, 8, 9]. Thus, the impact of overloading on network robustness is an important consideration for 5G NOMA network design.

While NOMA is not limited to 5G networks, we focus on 5G New Radio (NR) given the body of 5G-related NOMA research [5, 6, 7, 10, 11] and ongoing discussion of NOMA in the 3rd Generation Partnership Project [12]. We find specific motivation for the study of NOMA robustness in the context of 5G NR aerial access networks (AAN) [13]. AANs can provide 5G service for underserved areas and disaster recovery/public safety via unmanned aerial vehicle (UAV) base stations [14]. Dense network deployments are less likely in these scenarios making the sub-6 GHz range particularly important. Though much emphasis is placed on massive multiple input multiple output (mMIMO) to achieve 5G performance [2], the spatial multiplexing gains of mMIMO are less feasible in sub-6 GHz AANs due to a combination of increased channel correlation for LOS links [15] and the size, weight, and power requirements of a mMIMO array [16]. Thus, NOMA is a more prominent candidate to support 5G performance objectives in sub-6 GHz AANs.

In this paper, we investigate the relationship between NOMA overloading and network robustness through temporal graph components. We develop a graph model and analytical framework to characterize temporal component membership. We extend our analysis to derive lower bounds and probability expressions for the number of medium access control (MAC) frames required to achieve strong temporal connectedness between all NDs. These measures of network robustness are relevant for real-time Internet of Things [17] and distributed computing applications [18, 19] with small message lengths in which RB allocation for timely message exchange is an important factor in application performance.

The remainder of this paper is organized as follows. Section 2 discusses related work. Section 3 introduces the concepts of temporal connectedness and

HįCSS

components. Section 4 introduces our NOMA graph model. Section 5 describes our stochastic temporal component framework followed by individual and joint temporal component analysis and Sections 6-7. We present our affine graph analysis and simulation results in Section 8, and conclude the paper in Section 9.

2.   Related work

The robustness of wireless networks is often approached through stochastic geometry (SG) analysis due to the interference-limited nature of OMA [20, 21, 22]. SG is a well-studied approach that provides robustness metrics such as outage probability and bit error probability through the spatial averaging of random point processes, and has been applied for NOMA analysis [23, 24]. Recent SG analysis also includes a temporal aspect through queuing theory to consider maximum traffic density and age of information [25, 26]. However, the concept of overloading reduces the primacy of interference-based analysis since NDs in NOMA networks interfere by design. Therefore, investigating NOMA network robustness through temporal graph component metrics is an appropriate complement to SG analysis.

Much of the NOMA literature investigates methods to achieve 5G performance objectives under practical constraints [4, 6, 7, 10, 27, 28, 29]. Within this body of work, several researchers have considered overloading through standard physical and network layer measures of robustness. The link level simulation results presented in [30] compare uplink NOMA to OFDMA by assessing the effect of variable overloading on block error rate at multiple signal-to-noise ratios. Stability conditions for uplink code-domain NOMA are derived in [28] where stability is defined in terms of the queue size of NDs attempting to transmit in each time slot. Similarly, overloading stability in different downlink power-domain NOMA scenarios is examined in [31] where stability is defined as the power allocation among NDs that is required for successful signal differentiation. Our contribution provides a NOMA graph model and analytical framework that measures the effect of variable overloading on temporal connectedness between NDs.

3.   Temporal connectedness and components

Network connectedness is a common measure of function and robustness in static networks [32, 33]. Measures of connectedness for static networks are extended to temporal (time-varying) networks in [34, 35]. Unlike a static network, two nodes in a temporal network may only be connected for some finite duration of time, called a contact. A time-ordered sequence of contacts that allows two nodes to reach each other without traversing an edge more than once is a temporal path. Thus, node i is *temporally connected* to node j if there exists a temporal path from i to j.

Temporal connectedness is not a symmetric relation (i.e., a temporal path from node i to j does not imply a temporal path from node j to i), giving rise to definitions of strong and weak connectedness similar to directed static networks. Two nodes i and j of a time-varying graph are strongly connected if i is temporally connected to j, and also j is temporally connected to i. Similarly, two nodes i and j of a time-varying graph are weakly connected if i is temporally connected to j, and j is also temporally connected to i in the underlying undirected time-varying graph. This sets up the following temporal component definitions [34, 35]:

- *Temporal Strongly Connected Components* (TSCC): A set of nodes of a time-varying graph in which each node of the set is temporally strongly connected to all other nodes in the set.
- *Temporal Weakly Connected Components* (TWCC): A set of nodes of a time-varying graph in which each node of the set is temporally weakly connected to all other nodes in the set.

Finally, the *affine graph* of a time-varying network is the underlying undirected static graph associated with a specified time interval such that two nodes i and j are connected in the affine graph only if they are strongly connected in the time-varying graph [34, 35].

4.   NOMA graph model

The relationship between NDs, RBs, and the access point (AP) in a NOMA wireless network can be represented by a graph with a combination of dependency and functional connectivity edges as depicted in Figure 1. The two graphs correspond to the time division duplex (TDD) uplink and downlink subframes of a single MAC frame. The NDs are dependent on the use of RBs to transmit (receive) information to (from) the AP. The dependency links show the use of a specific RB by a ND and the directed connectivity links show the transmission (reception) of information. The downlink subframe only contains directed connectivity links from the AP to the NDs and the uplink subframe only contains directed connectivity links from the NDs to the AP. The subframe graphs are individual realizations of the network dependency and connectivity links, and the TDD MAC frame represents the smallest time unit in which two different NDs can be
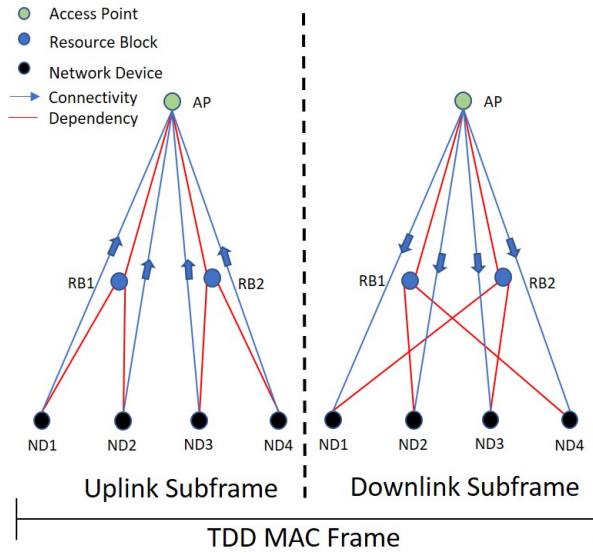
Figure 1. Dependency-Connectivity NOMA Graph Model.

bidirectionally connected.

In the context of 5G NR, this model directly corresponds to two consecutive slots with 15 kHz subcarrier spacing (i.e., the subframe and slot length are equal) in which all symbols are configured for uplink in the first slot and downlink in the second. However, the frame and subframe are general model constructs and can be adapted to any 5G NR subcarrier numerology.

4.1. Temporal network ensemble

The critical aspect in our model is the dependency relationship between the NDs and the RBs. Given that all the RBs are mapped to a single AP and the directed connectivity can be inferred from the subframe in which the dependency links occur, the mixed dependency-connectivity graphs in Figure 1 can be reduced to a bipartite graph $G(U, V, E)$ where U is the node set of RBs, V is the node set of NDs, and E is the edge set representing the allocation of RBs to NDs in each subframe.

We assume all RBs are allocated to NDs in each subframe and let the overloading ratio $z \in \mathbb{Z}^+$ and quantity of RBs $|U| \in \mathbb{Z}^+$ both be positive integers. Further, we let $|V| > z|U|$, deg $u = z$, $\forall u \in U$, and deg $v \in \{0, 1\}$, $\forall v \in V$. This representation allows each ND to potentially form an edge with no more than a single RB, and models the overloading ratio in the degree of each RB node. Since the degree of all RB nodes is the same and the degree of all ND nodes is either one or zero, the network can be specified by a labeled binary degree sequence $\mathbf{d}_v$ (degree vector) that indicates which NDs have received a RB allocation. Thus, the ensemble of possible network configurations is defined by the ensemble of all unique degree vectors of the ND node set parameterized by the number of NDs, RBs, and the overloading ratio z. We randomly and uniformly select a degree vector from the ensemble with replacement to generate a graph realization for a NOMA wireless network subframe. This sampling process is repeated until a temporal network with the required number of subframes is generated. The quantity of degree vectors that compose the temporal network ensemble is given by the binomial coefficient,

$$\binom{m}{zn} = \frac{m!}{zn!(m - zn)!}, \tag{1}$$

where $m = |V|$ is the number of NDs, $n = |U|$ is the number of RBs, and z is the overloading ratio. Additionally, we note that setting the maximum degree of ND nodes equal to one abstracts the implementation of some code-domain NOMA approaches, but still characterizes the aggregate allocation of RBs to NDs.

4.2. Relationship to existing graph models

Our proposed NOMA network representation is a hybrid adaptation of the well-known configuration model for static networks [36], and activity-driven model for temporal networks [37].

The configuration model is a random graph model that generates an ensemble of graphs from a set of nodes and a given degree sequence. The degree sequence is instantiated in the nodes by assigning a number of "stubs" equal to the degree of each node in the degree sequence. Two stubs are then chosen uniformly at random to connect and form an edge. The network ensemble is comprised of all possible stub "matchings," or the different ways in which these stubs can be connected [32, 33, 36]. However, this ensemble does not suit our purpose due to the unique nature of the explicit dependency and implied connectivity model. Once the stubs are assigned to NDs and RBs, we understand which NDs will have connectivity during that subframe. The number of different ways in which the ND stubs can be mapped to the RB stubs does not provide any additional information about the connectivity. Our model considers whether NDs can map to *any* RB by defining the ensemble as all possible matchings of stubs to nodes rather than all possible ways of connecting stubs that have already been assigned to nodes.

The activity-driven model is a snapshot representation of a temporal network that assigns an activity potential drawn from a probability distribution

to each node in each time step. The activity potential defines a probability that a node forms m undirected edges with m uniformly and randomly selected nodes [37, 38, 39]. Nodes which are inactive in a snapshot may still form an edge with another node if they are randomly selected for edge creation by an active node. The probability of RB allocation to each ND (discussed in Section 5.1) is similar to the activity potential, but remains static in each snapshot since it is defined by the NOMA network parameters. Additionally, each ND node that receives a RB allocation can form an edge with any RB node, but this is a dependency relationship. All functional connectivity edges terminate with the AP.

5. Stochastic temporal component framework

In this section, we develop a stochastic framework to evaluate temporal component membership.

5.1. Network sequence as a Bernoulli process

Allocating RBs to any individual ND in a graph realization generated from a degree vector drawn from the temporal network ensemble can be modeled as a Bernoulli random variable K with probability mass function (PMF) given by

$$\mathrm{f}_K[k] = \begin{cases} p & k = 1 \\ 1 - p & k = 0 \\ 0 & \text{otherwise} \end{cases}, \qquad (2)$$

where

$$p = \frac{zn}{m}. \qquad (3)$$

As discussed in Section 4.1, the temporal NOMA wireless network is generated from random uniform sampling of the network ensemble. This means the RB allocation is an independent Bernoulli trial in each graph realization for a single ND. Thus, a sequence of N binary degree vectors randomly and uniformly sampled from the network ensemble constitute a Bernoulli random process with N realizations occurring across all labeled nodes $v \in V$ with a probability mass function for each realization given by Equation 2. In matrix form, this is an $N \times m$ matrix, where the rows are the randomly selected degree vectors in ascending temporal order from 1 to N, the columns are the labeled nodes of V, and the ij^{th} elements are the degree of node v_j in network realization t_i. Thus, each column represents the Bernoulli random process of each labeled node. A visual representation is shown in Figure 2. The green rectangle highlights a single network realization occurring at t_i,

and the blue rectangle highlights a Bernoulli process of RB allocation for a single node, v_j.

Figure 2. Network Sequence Matrix

5.2. Probability of resource block allocation

Characterizing the network sequence of each ND as a Bernoulli random process provides a basis to consider the probability that a ND receives k RB allocations in a network sequence of N realizations. The statistical independence of each RB allocation event means the joint probability of any specific RB allocation sequence is the product all independent probabilities. If the specific order of the RB allocations is not relevant, then we also multiply the joint probability by the number of combinations that could result from a specific number of RB allocations and "nulls." This combined product is the well-known binomial distribution. Thus, the probability of k RB allocations in a network sequence of N realizations is a binomial random variable with probability mass function

$$\mathrm{f}_K[k] = \begin{cases} \binom{N}{k} p^k (1-p)^{N-k} & 0 \le k \le N \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

where p is given by Equation 3. This distribution can also be used to determine the probability that the number of RB allocations falls within a range that is bounded by network sequence length, and can be approximated by a Gaussian distribution as N becomes large [40].

Though our RB allocation model is a departure from channel-based dynamic scheduling, the Bernoulli random process ensures each ND receives a nearly equal number of RB allocations as N becomes large [40]. Thus, our approach provides a maximum fairness baseline for RB allocation that is applicable to narrow band transmissions in which minimum data rate requirements are met for all NDs [8].

5.3. Temporal component membership

Network device membership in a temporal component can be defined in terms of connectivity

with the AP, or connectivity with the other NDs in the network. We focus on the latter definition, and treat the AP as a relay rather than a gateway as might be the case in 5G AANs [13]. This requires a modification to the definition of weak connectedness in [34, 35] because the network topology renders the TWCC indistinguishable from the TSCC. Thus, we adopt the temporal path definition of [34, 35], but amend the definition of weak temporal connectedness as follows:

Weak Temporal Connectedness: two nodes i and j of a time-varying graph are weakly connected if, either i is temporally connected to j, *or* j is temporally connected to i, in the underlying undirected time-varying graph.

The TWCC definition remains as stated in Section 3.

6. Individual probability of temporal component membership

Defining temporal component membership based on connectivity between NDs through the AP requires two subframes to evaluate which NDs belong to which temporal components. We assume all NOMA wireless networks of interest are parameterized by a sufficient combination of RBs and overloading ratio to support RB allocation to more than one ND in a subframe, and correspondingly define the temporal component event tree in Figure 3.
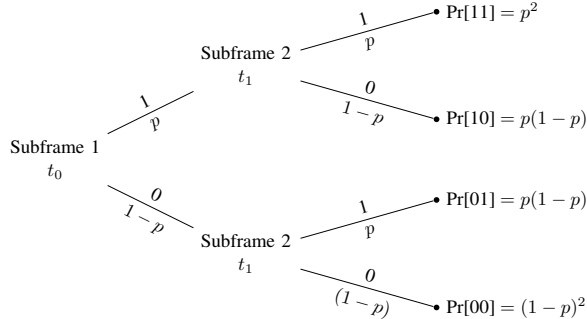

Figure 3.   Temporal Component Event Tree

This event tree shows the temporal RB allocation sample space for a single ND over two subframes, t_0 and t_1 (uplink and downlink). Resource block allocation is designated by the value 1 with probability p, and a lack of RB allocation (null) is designated by a 0 with probability $1 - p$. Two subframes yield a binary sequence of length two with four possible outcomes, $\{11, 10, 01, 00\}$. Strong temporal connectedness requires bidirectional temporal paths between two NDs in the directed graph, which corresponds to a RB allocation in both subframes, e.g. $\{11\}$. Weak temporal connectedness requires a unidirectional temporal path between two NDs in

the underlying undirected time-varying graph which corresponds to one RB allocation in either subframe, either $\{10\}$ or $\{01\}$. Finally, an isolated node corresponds to zero RB allocations in either subframe, $\{00\}$. These mutually exclusive and exhaustive event probabilities are the result of an underlying independent identically distributed (IID) random process; thus, we use each event probability and its complement to define a new Bernoulli random process that characterizes the individual ND membership probability for the corresponding temporal component.

Let the probability that any ND $u \in U$ is a member of the TSCC, TWCC (only), or isolated during any frame in a network sequence be denoted by p_s, p_w, and p_i, respectively, where

$$p_s = p^2, \ p_w = 2p(1-p), \ p_i = (1-p)^2, \ \text{and}$$

$$p_s + p_w + p_i = 1.$$

The probability that a ND is a member of the TSCC in any arbitrary frame is given by substituting p_s for p in Equation 2. Similarly, the probability that a ND is a member of the TSCC for k frames in a network sequence of N frames is given by substituting p_s for p in Equation 4. This same approach also applies to the event probabilities for the TWCC and isolated events.

7. Joint probability of temporal component membership

The analysis of joint ND temporal component membership is different than for an individual ND. Recall from Section 5.1 that we can analyze one ND over an arbitrary number of network realizations from $\{1, ..., N\}$, where each realization is an IID Bernoulli random variable, and the joint probability of any specific sequence of events is the product of the individual event probabilities. However, each network realization is the result of a degree vector with fixed values dependent on the network parameters, so the probability of RB allocation for two or more NDs are not independent within a *single* network realization.

The binary characterization of RB allocation facilitates use of the Hypergeometric distribution for joint RB probability analysis within a single network realization. The distribution defines the probability of drawing a sample of n items containing k successes without replacement from a population N with a known number of successes M [40]. The PMF is given by

$$f_K[k] = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}. \tag{5}$$

In our NOMA network model, the successes are those NDs in the sample that receive a RB allocation. Substituting the NOMA parameters into Equation 5 gives

$$\mathrm{f}_K[k] = \frac{\binom{nz}{k}\binom{m-nz}{m_s-k}}{\binom{m}{m_s}}, \qquad (6)$$

where m_s is the number of NDs sampled, m is the total number of NDs in the network, the quantity of known successes is the number of RBs n multiplied by the overloading ratio z, and k is the number of RBs allocated to the unlabeled NDs in the sample.

Determining the joint probability of RB allocation among any *labeled* combination of NDs requires equally distributing the unlabeled joint probability over the number of possible unique RB allocations to the labeled NDs, where the number of possible RB combinations is given by the binomial coefficient $\binom{m_s}{k}$. Thus, the labeled probabilities are given by

$$\mathrm{f}_K[k]_{\text{labeled}} = \frac{\mathrm{f}_K[k]}{\binom{m_s}{k}}. \qquad (7)$$

Once the joint probability has been determined, it is fixed over all frames in the network sequence and each frame is an IID realization. Thus, we can define an exhaustive and mutually exclusive joint probability space with the desired number of NDs and calculate temporal component membership over an arbitrary number of network realizations using the same approach discussed in Section 6.

8. Affine graph analysis

In this section, we consider the minimum number of MAC frames required to achieve a complete affine graph and the probability of occurrence. Recall that the affine graph provides a temporal measure of bidirectional connectivity between ND pairs in the network. Pairwise ND communication is central in distributed computing; thus, this analysis relates NOMA overloading to the supported convergence time of distributed computing algorithms that could be deployed on 5G AANs [18, 19].

Despite research that shows more efficient wireless communications methods using downlink broadcast for distributed computation [41], our analysis considers individual scheduling for NDs in the downlink subframe. This is consistent with the 5G NR MAC sublayer logical to transport channel mapping. All user-plane traffic is mapped from the dedicated traffic channel (logical) to the downlink shared channel (transport), and not to a broadcast channel which is reserved for system information [2, 3].

8.1. Mathematical framework

Our mathematical framework for affine graph analysis includes connectivity potential matrices, a frame connectivity matrix, and a queue matrix.

Connectivity potential matrices translate the allocation of RBs into the corresponding potential connectivity they enable between NDs.

Uplink Potential Matrix: Given a binary RB allocation vector $\mathbf{d}_v = [d_1, ..., d_m]$ for m NDs, let $\mathbf{P}_{UL}$ be the uplink potential matrix given by

$$\mathbf{P}_{UL} = (\mathbf{1} - \mathbf{I}) \odot \mathbf{d}^T, \qquad (8)$$

where $\mathbf{1}$ is $m \times m$ uniform matrix, $\mathbf{I}$ is $m \times m$ identity matrix, and $\odot$ denotes the Hadamard Product.

Downlink Potential Matrix: Given a binary RB allocation vector $\mathbf{d}_v = [d_1, ..., d_m]$ for m NDs, let $\mathbf{P}_{DL}$ be the downlink potential matrix given by

$$\mathbf{P}_{DL} = (\mathbf{1} - \mathbf{I}) \odot \mathbf{d}. \qquad (9)$$

Frame Connectivity Matrix: The frame connectivity matrix $\mathbf{F}$ is a directed adjacency matrix that defines the connectivity achieved between NDs during a single frame as a result of the uplink and downlink RB allocations. The frame connectivity matrix is the Hadamard Product, of the uplink and downlink potential matrices, $\mathbf{P}_{UL}$ and $\mathbf{P}_{DL}$, given by

$$\mathbf{F} = \mathbf{P}_{UL} \odot \mathbf{P}_{DL}. \qquad (10)$$

Frame Queue Matrix: The frame connectivity matrix defines the achieved connectivity in each frame, but does not account for "undelivered" messages that result from the ijth elements of the uplink potential matrix that go unmatched with the downlink potential matrix in the same frame. This information must be captured since a pair of NDs that were unable to communicate over the time interval of one frame might be able to communicate if the time interval is extended to two frames. A queue matrix to capture this information corresponds to buffers in wireless networks where packets are queued until they can be transmitted. This matrix will change during each frame depending on the uplink and downlink RB allocations, so we define the queue matrix $\mathbf{Q}$ recursively as

$$\mathbf{Q}_t = \mathbf{P}_{UL} + \mathbf{Q}_{t-1} - \mathbf{F}_t, \qquad (11)$$

where $\mathbf{Q}_0$ is the $m \times m$ null matrix $\mathbf{0}$. The queue matrix that feeds into the next frame is the sum of the uplink potential matrix and previous queue matrix, minus the achieved transmissions resulting from the frame connectivity matrix.

Dynamic Frame Connectivity Matrix: The time-dependent definition of the frame queue matrix $\mathbf{Q}_t$ implies the frame connectivity matrix $\mathbf{F}$ must account for queued messages from the previous frame. We define the frame connectivity matrix at time t

$$\mathbf{F}_t = \mathbf{1}_{\mathbb{Z}^+}(\mathbf{P}_{UL} + \mathbf{Q}_{t-1}) \odot \mathbf{P}_{DL}, \qquad (12)$$

where $\mathbf{1}_{\mathbb{Z}^+}(\mathbf{X})$ is the indicator function for the set of positive integers to generate a binary matrix. This conceptualization captures the directed connectivity between nodes that occurs over more than one frame.

8.2. Affine graph

Each dynamic frame connectivity matrix captures the achieved connectivity in frame t. However, the affine graph represents strong temporal connectedness between node pairs over an observation interval. Thus, we must sum all the individual frame connectivity matrices to produce an aggregate picture of network connectivity during an arbitrary observation interval $t = [T-l, T-l+1, ..., T-1, T]$ selected from the network sequence. If we let $l = T+1$, then t is the entire network sequence, and the static projection directed adjacency matrix of the NOMA wireless network is given by

$$\mathbf{S} = \sum_{t=1}^{T} \mathbf{F}_t. \qquad (13)$$

Since $\mathbf{S}$ represents all achieved connectivity over the entire network sequence, we calculate the affine graph

$$\mathbf{A} = \mathbf{1}_{\mathbb{Z}^+}(\mathbf{S} \odot \mathbf{S}^T) \qquad (14)$$

Recall that the ith row of $\mathbf{S}$ represents the achieved directional connectivity from ND i to ND j ($\forall j$ with $i \neq j$), and the jth column of $\mathbf{S}$ represents the achieved directional connectivity from ND j to ND i. Thus, taking the Hadamard Product of $\mathbf{S}$ with its transpose only yields a non-zero value for the ijth element of $\mathbf{A}$ if bidirectional connectivity has been established between nodes i and j. Applying the indicator function results in an unweighted binary matrix.

8.3. Analytical results

The minimum number of frames required to form a complete affine graph is the sum of the number of frames required for all NDs to receive an uplink RB allocation, and the number of frames required for all NDs to receive a downlink RB allocation after all have received an uplink RB allocation. The corresponding probability of this event depends on the network parameters m,

n, and z. We now present the possible cases for this relationship.

Case 1: $zn = m - 1$. This case occurs when all but one ND can receive a RB allocation in each subframe. A complete affine graph can be achieved in two frames with probability

$$\Pr[F_{min}] = p^2(1 - p), \qquad (15)$$

where p is the probability of RB allocation for an individual ND defined in Equation 3.

Consider the case in which the single ND i that does not receive an uplink RB allocation in the first subframe does receive a RB allocation in the first downlink subframe. This means ND i has received from all other NDs (since ND i does not transmit to itself). If ND i receives an uplink RB allocation in the second uplink subframe, then all NDs have received an uplink RB allocation. Since ND i has already received from all other NDs, it does not require an additional downlink RB allocation, so if all other $m - 1$ NDs receive a downlink RB allocation in the second downlink subframe, then a complete affine graph will have been achieved. This sequence of frames can be characterized entirely by the behavior of a single ND i, with probability of RB allocation, p.

Case 2: $\frac{m}{2} \leq zn < m - 1$. This case occurs when at least half of the NDs in the network can receive a RB allocation in each subframe, but less than ND $- 1$. A complete affine graph can be achieved in three frames with probability

$$\Pr[F_{min}] = \left[\frac{\binom{zn}{2zn-m}}{\binom{m}{zn}} \right]^2. \qquad (16)$$

In this case, a minimum of two uplink subframes are required for all NDs to receive at least one uplink RB allocation. Similarly, a minimum of two downlink subframes are required for all NDs to receive at least one downlink RB allocation. All NDs must receive a downlink RB allocation *after* all NDs have received an uplink RB allocation in order to receive a transmission directly in the same frame, or indirectly from the message queue. Thus, the downlink RB allocation must begin in the second downlink subframe, and conclude in the third downlink subframe.

Consider a NOMA wireless network parameterized by $m = 6$, $n = 2$, and $z = 2$. In the first frame, the probability of selecting four NDs which have not yet received an uplink RB allocation is equal to one since the network sequence has just begun. In the second frame, we must select the $m - zn = 2$ NDs that did not receive an uplink RB allocation in the first uplink

subframe, as well as $zn - (m - zn) = 2$ of the $zn = 4$ NDs that did receive an uplink RB allocation in the first uplink subframe. The probability of this event is

$$\frac{\binom{m-zn}{m-zn}\binom{zn}{zn-(m-zn)}}{\binom{m}{zn}} = \frac{\binom{zn}{2zn-m}}{\binom{m}{zn}} = \frac{\binom{2}{2}\binom{4}{2}}{\binom{6}{4}} = \frac{\binom{4}{2}}{\binom{6}{4}}.$$

The second downlink subframe is the first opportunity for NDs to receive a downlink RB allocation after all NDs have received an uplink RB allocation. Thus, the RB allocation sequence that was required for the uplink in the first and second frames must repeat for the downlink in the second and third frames to achieve a complete affine graph, which gives

$$\Pr[F_{min}] = \left[\frac{\binom{4}{2}}{\binom{6}{4}} \right]^2.$$

Generalizing this to any combination of NDs, RBs, and overloading ratio that meet these conditions results in Equation 16. Note that when $zn = \frac{m}{2}$, Equation 16 reduces to

$$\Pr[F_{min}] = \binom{m}{m/2}^{-2}. \tag{17}$$

Case 3: $1 < zn < \frac{m}{2}$. This case occurs when more than one, but less than half, of the NDs can receive a RB allocation in each subframe. There are three sub-cases within Case 3.

Subcase 3.1: $m \mod zn = 0$. This occurs when the product of RBs and the overloading ratio divide evenly into the number of NDs. The number of frames required for all NDs to receive a RB allocation is equal to $\frac{ND}{zn}$, and the downlink RB allocation can begin in the same frame that all NDs receive an uplink allocation. Hence, the minimum number of frames required to achieve a complete affine graph is given by

$$F_{min} = \frac{2m}{zn} - 1. \tag{18}$$

The probability of this event is a product of the Hypergeometric distribution evaluated with constant population (m), sample size (zn), and desired number of successes (zn), but a decreasing number of known successes in the population. We treat the NDs that have yet to receive an uplink or downlink RB allocation as the number of successes in the population, and decrement this value by zn after each subframe. The probability of selecting zn NDs that have not yet received a RB allocation in the first uplink subframe is one since no NDs have received an RB allocation at the beginning of

the network sequence. In the second uplink subframe, we must select zn of $m - zn$ NDs that have not yet received an uplink RB allocation. This decrementing operation continues until all NDs have received an uplink RB allocation, and then repeats for the downlink, resulting in a probability given by

$$\Pr[F_{min}] = \prod_{i=1}^{\frac{m}{zn}-1} \frac{\binom{m-i(zn)}{zn}^2}{\binom{m}{zn}^2}. \tag{19}$$

Since we never select any NDs that have already received a RB allocation, the second term in the numerator from Equation 6 reduces to one and the square operation accounts for the repetition of the process in the uplink and downlink.

Subcase 3.2: $m \mod zn = 1$. This occurs when any multiple of the product of RBs and the overloading ratio is one less than the number of NDs. The logic for the minimum frames required for a complete affine graph is the same as Case 1, except there are more frames prior to reaching the frame in which only a single ND has not yet received an uplink RB allocation. Accounting for those additional frames, the minimum number of frames required to achieve a complete affine graph is given by

$$F_{min} = 2\left\lfloor \frac{m}{zn} \right\rfloor. \tag{20}$$

The probability formulation for this case is similar to Subcase 3.1 in that we take a product of Hypergeometric distributions evaluated at decrementing values of the number of successes in the population. The primary differences occur in downlink subframe $\left\lfloor \frac{m}{zn} \right\rfloor$ and in uplink subframe $\left\lfloor \frac{m}{zn} \right\rfloor + 1$ which immediately follows. In downlink subframe $\left\lfloor \frac{m}{zn} \right\rfloor$, all NDs have received an uplink RB allocation except ND i. Thus, ND i must receive a downlink RB allocation in this subframe to have received from all other NDs. This occurs with probability $\binom{m-1}{zn-1}/\binom{m}{zn}$. Similarly, ND i must receive an uplink RB allocation in the following uplink subframe so all remaining $m - 1$ NDs can receive from ND i in the following $\left\lfloor \frac{m}{zn} \right\rfloor$ subframes. This also occurs with probability $\binom{m-1}{zn-1}/\binom{m}{zn}$. Thus, the total probability of achieving a complete affine graph in this subcase is given by

$$\Pr[F_{min}] = \frac{\binom{m-1}{zn-1}^2}{\binom{m}{zn}^{2\alpha+1}} \prod_{i=1}^{\alpha-1} \binom{m-izn}{zn} \\ \prod_{j=0}^{\alpha-1} \binom{m-1-jzn}{zn}, \tag{21}$$

where we have collected all denominators of the product into the denominator of the first term, the second and third terms are the decrementing Hypergeometric distributions for the uplink and downlink, respectively, and $\alpha = \left\lfloor \frac{m}{zn} \right\rfloor$.

Subcase 3.3: $m \mod zn > 1$. This occurs when any multiple of the product of RBs and the overloading ratio does not fall in either Subcase 3.1 or 3.2. The minimum number of frames required to achieve a complete affine graph is given by

$$F_{min} = 2 \left\lceil \frac{m}{zn} \right\rceil - 1. \tag{22}$$

The probability of achieving the affine graph in the minimum number of frames can be directly calculated, but does not have a concise expression. Rather, we must determine all possible ways in which all NDs can receive a RB allocation in the minimum number of frames, determine the probability of each of these possibilities, and then sum the pairwise multiplication of all probabilities to account for all possible uplink and downlink combinations of RB allocations.

8.4. Simulation results

Simulations to test our analytical results were conducted in MATLAB. For each case with an expression for $\Pr[F_{min}]$ we defined a number of NDs (m), RBs (n), and overloading ratio (z) that met the case conditions and generated the corresponding temporal network ensemble. Next, we generated a network sequence of length N through random uniform sampling from the ensemble where N was set equal to F_{min} based on the appropriate case. Finally, we calculated the resulting affine graph from each network sequence using the mathematical framework of Sections 8.1 and 8.2.

We generated 5×10^6 network sequences for each case and recorded the number of complete affine graphs achieved. We compared the simulation result with the predicted value for each case and calculated the absolute percent error. The cases, network parameters, and error results are shown in Table 1.

Table 1. Simulation Results

Case	ND	RB	z	Error (%)
1	9	2	4	0.1341
2	10	2	4	0.0574
2	10	2	2.5	0.9338
3.1	6	2	1	0.2125
3.2	9	2	2	0.1117

The largest absolute percent error across all cases is 0.9338% which lends strong support for the accuracy of the analytical results. We note this error corresponds to the lower bound of Case 2 in which exactly half of the NDs can receive a RB allocation in each frame. This relationship between m and zn corresponds to the maximum size of the network ensemble described in Equation 1, so we expect a larger error than other simulation cases with a smaller network ensemble.

9. Conclusion

In this paper, we explored the relationship between NOMA overloading and network robustness through temporal connectedness in the context of a 5G AAN. We proposed a mixed dependency-connectivity graph model for a NOMA network, developed a stochastic temporal component framework, and characterized the probability of temporal component membership in terms of the overloading ratio. We extended this analysis to the affine graph by developing analytical expressions for the minimum number of MAC frames in which bidirectional connectivity can be achieved between all NDs, and the probability of these events. Finally, we supported our analytical results through simulation.

References

[1] M. Series, "IMT Vision–Framework and overall objectives of the future development of IMT for 2020 and beyond," *Recommendation ITU*, vol. 2083, 2015.

[2] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The Next Generation Wireless Access Technology*. Academic Press, 2018.

[3] S. Ahmadi, *5G NR: Architecture, Technology, Implementation, and Operation of 3GPP New Radio Standards*. Academic Press, 2019.

[4] Y. Chen, A. Bayesteh, Y. Wu, B. Ren, S. Kang, S. Sun, Q. Xiong, C. Qian, B. Yu, Z. Ding, *et al.*, "Toward the standardization of non-orthogonal multiple access for next generation wireless networks," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 19–27, 2018.

[5] S. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 721–742, 2016.

[6] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE communications surveys & tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018.

[7] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple access techniques for 5G wireless networks and beyond*. Springer, 2019.

[8] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1805–1838, 2020.

[9] A. E. Mostafa, Y. Zhou, and V. W. Wong, "Connection density maximization of narrowband IoT systems with NOMA," *IEEE Transactions on Wireless Communications*, vol. 18, no. 10, pp. 4708–4722, 2019.

[10] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6010–6023, 2015.

[11] Y. Yuan, Z. Yuan, and L. Tian, "5G non-orthogonal multiple access study in 3GPP," *IEEE Communications Magazine*, vol. 58, no. 7, pp. 90–96, 2020.

[12] 3GPP, "Release 16: Study on Non-Orthogonal Multiple Access (NOMA) for NR," Technical Report (TR) 38.812, 3rd Generation Partnership Project (3GPP), Dec 2018. Version 16.0.0.

[13] 3GPP, "Release 15: Study on New Radio (NR) to support non-terrestrial networks," Technical Report (TR) 38.811, 3rd Generation Partnership Project (3GPP), Sep 2020. Version 15.4.0.

[14] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE communications surveys & tutorials*, vol. 21, no. 3, pp. 2334–2360, 2019.

[15] A. A. Nasir, H. D. Tuan, T. Q. Duong, and H. V. Poor, "UAV-enabled communication using NOMA," *IEEE Transactions on Communications*, vol. 67, no. 7, pp. 5126–5138, 2019.

[16] E. Bjornson, L. Van der Perre, S. Buzzi, and E. G. Larsson, "Massive mimo in sub-6 ghz and mmwave: Physical, practical, and use-case differences," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 100–108, 2019.

[17] T. Park, W. Saad, and B. Zhou, "Centralized and distributed age of information minimization with nonlinear aging functions in the internet of things," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 8437–8455, 2020.

[18] A. Nedić, A. Olshevsky, and M. G. Rabbat, "Network topology and communication-computation tradeoffs in decentralized optimization," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 953–976, 2018.

[19] Y. Shi, K. Yang, T. Jiang, J. Zhang, and K. B. Letaief, "Communication-efficient edge AI: Algorithms and systems," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2167–2191, 2020.

[20] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE journal on selected areas in communications*, vol. 27, no. 7, pp. 1029–1046, 2009.

[21] M. Z. Win, P. C. Pinto, and L. A. Shepp, "A mathematical theory of network interference and its applications," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 205–230, 2009.

[22] H. ElSawy, A. Sultan-Salem, M.-S. Alouini, and M. Z. Win, "Modeling and analysis of cellular networks using stochastic geometry: A tutorial," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 167–203, 2016.

[23] K. S. Ali, M. Haenggi, H. ElSawy, A. Chaaban, and M.-S. Alouini, "Downlink non-orthogonal multiple access (noma) in poisson networks," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1613–1628, 2018.

[24] X. Yue, Z. Qin, Y. Liu, S. Kang, and Y. Chen, "A unified framework for non-orthogonal multiple access," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5346–5359, 2018.

[25] G. Chisci, H. ElSawy, A. Conti, M.-S. Alouini, and M. Z. Win, "Uncoordinated massive wireless networks: Spatiotemporal models and multiaccess strategies," *IEEE/ACM Transactions on Networking*, vol. 27, no. 3, pp. 918–931, 2019.

[26] M. Emara, H. ElSawy, and G. Bauch, "A spatiotemporal model for peak aoi in uplink IoT networks: Time versus event-triggered traffic," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6762–6777, 2020.

[27] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE signal processing letters*, vol. 21, no. 12, pp. 1501–1505, 2014.

[28] M. Shirvanimoghaddam, M. Condoluci, M. Dohler, and S. J. Johnson, "On the fundamental limits of random non-orthogonal multiple access in cellular massive IoT," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2238–2252, 2017.

[29] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, "A survey of NOMA: Current status and open research challenges," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 179–189, 2020.

[30] HiSilicon, Huawei, "LLS results for uplink multiple access," Tech. Rep. R1-164037, 3rd Generation Partnership Project (3GPP), May 2016.

[31] Y. Huang, J. Wang, and J. Zhu, "Optimal power allocation for downlink NOMA systems," in *Multiple Access Techniques for 5G Wireless Networks and Beyond*, pp. 195–227, Springer, 2019.

[32] G. Bianconi, *Multilayer networks: structure and function*. Oxford university press, 2018.

[33] M. Newman, *Networks*. Oxford university press, 2018.

[34] V. Nicosia, J. Tang, M. Musolesi, G. Russo, C. Mascolo, and V. Latora, "Components in time-varying graphs," *Chaos: An interdisciplinary journal of nonlinear science*, vol. 22, no. 2, p. 023101, 2012.

[35] V. Nicosia, J. Tang, C. Mascolo, M. Musolesi, G. Russo, and V. Latora, "Graph metrics for temporal networks," in *Temporal networks*, pp. 15–40, Springer, 2013.

[36] B. Bollobás, "A probabilistic proof of an asymptotic formula for the number of labelled regular graphs," *European Journal of Combinatorics*, vol. 1, no. 4, pp. 311–316, 1980.

[37] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, "Activity driven modeling of time varying networks," *Scientific reports*, vol. 2, no. 1, pp. 1–7, 2012.

[38] P. Holme, "Modern temporal network theory: a colloquium," *The European Physical Journal B*, vol. 88, no. 9, pp. 1–30, 2015.

[39] N. Masuda and R. Lambiotte, *Guide To Temporal Networks, A*, vol. 6. World Scientific, 2020.

[40] A. Papoulis and U. Pillai, *Probability, random variables, and stochastic processes*. Boston: McGraw-Hill, 4th ed. ed., 2002.

[41] S. Li, Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "A scalable framework for wireless distributed computing," *IEEE/ACM Transactions on Networking*, vol. 25, no. 5, pp. 2643–2654, 2017.