

# A Note on the Buffer Overlap Among Nodes Performing Random Network Coding in Wireless Ad Hoc Networks

Riccardo Masiero<sup>†</sup>, Daniele Munaretto<sup>\*</sup>, Michele Rossi<sup>†</sup>, Joerg Widmer<sup>\*</sup> and Michele Zorzi<sup>†</sup>

<sup>†</sup>DEI, University of Padova, via Gradenigo 6/B – 35131, Padova, Italy

<sup>\*</sup>DoCoMo Euro-Labs, Landsberger Strasse 312 – 80687 Munich, Germany.

**Abstract**—Network coding is a technique which is particularly suitable for the dissemination of data in distributed ad hoc networks. The definition of a mathematical model that describes the interactions among nodes and, in particular, their relationship in terms of buffer subspaces is still an open and challenging problem. The contribution of this paper is an analysis of the relationship between the network topology and the subspace overlap among nodes. This analysis can be used to establish criteria for the design of packet combination policies in diverse networking scenarios. Differently from previous studies, we will explicitly take the overlap among subspaces into account through a framework comprising networks with fixed as well as mobile nodes.

## I. INTRODUCTION

Efficient data delivery is of very high importance for distributed wireless networking. Random Linear Network Coding (RLNC) [1] is a technique that is particularly suitable for the dissemination of data as it can be used in a distributed and completely unsynchronized manner. Furthermore, network coding algorithms can exploit the broadcast nature of the wireless medium to boost performance. The random mixing of different data flows makes data dissemination robust, which is particularly important in, e.g., mobile ad hoc and sensor networks where node failures are common. Even though previous work dealt with both theoretical and practical schemes for RLNC, it is still unclear how packets should be combined in order to get the highest benefits in terms of throughput, delay, energy efficiency, and data persistence [2] (the amount of information that can be decoded at the data gathering point(s) at any given time).

A mathematical model that describes the interactions among nodes performing network coding is a powerful tool from a protocol-design perspective. However, defining such a model is a challenging problem. With RLNC, a network node sends out random vectors from the information space spanned by the packets received thus far. However, it is also possible to create packets from only a subspace of the whole information space available at that node. The dimension of this subspace has an impact on the encoding and decoding complexity as well as the efficiency of the data dissemination process. Our aim is to describe the RLNC transmission dynamics so as to predict, with sufficient accuracy, the evolution of the dissemination process. To do this, we model the data delivery using combinatoric tools that allow us to track the overlap between the sub-spaces spanned by the buffers of different users.

The paper is organized as follows. In section II we overview the related work. In section III we illustrate two mathematical models to describe the buffer overlap among nodes performing RLNC and we establish a relation between buffer overlap and the probability that the transmission of a new packet, coded using RLNC, provides innovative information at the receiver. In sections IV and V we present simulation scenarios and results, respectively. Section VI concludes the paper.

## II. RELATED WORK

Several studies have been carried out to understand the dynamics of RLNC in distributed networks. As an example, an analysis related to our own can be found in [3] and [4], where the authors exploit the properties of the subspaces spanned by the collected information vectors to identify the topological structure of the underlying network graph. In [4] the subspace observation is further used for topology management, to avoid bottlenecks and clustering in network-coded peer-to-peer systems. In [5] it has been shown that coding generally provides benefits in terms of data persistence. Nevertheless, always combining the entire available information when sending a new data packet might leave coded packets undecodable (e.g., after a network failure), thus reducing the data recovery performance. References [2], [6] and [7] address this issue by proposing code degree distributions that maximize data persistence. [2] presents a new class of codes called Growth Codes, that have been generalized in [6]. These papers demonstrate that optimal combination policies exist, even though the analysis is based on the assumption that the information subspaces available at the nodes are uncorrelated. This is true for information exchange at random encounters among nodes or for very high mobility. However, it does not well capture protocol behavior in realistic networks (especially in the presence of moderate mobility or static networks), as shown in [7]. In this last paper, the authors show that networks with a connectivity graph that changes significantly between subsequent transmissions are only representative of a small class of realistic networks. These models do not respect the dynamics of the underlying connectivity structure of many networks and this may seriously impact the performance of the techniques of [2]. The coding rules proposed in [7] are based on heuristics; our aim in this paper is to go beyond this by analyzing buffer dynamics through a mathematical model.

## III. THEORETICAL MODELS

Consider a network with  $N$  nodes performing data dissemination with network coding over a finite field  $\mathbb{F}_q$ . At node  $i$ , the incoming packets containing the information vectors received up to time  $t$  form a matrix  $Y_i^t$ . Let  $\langle Y_i^t \rangle$  denote the subspace spanned by the rows of  $Y_i^t$ . With RLNC, at transmission time  $t$ , the node sends an encoded packet containing a linear combination  $y_i^t = mY_i^t$  where  $m$  is a local encoding vector of random coefficients in  $\mathbb{F}_q$  [8]. (For ease of notation, we will omit index  $t$  in the remainder of the paper.)

Instead, the node may also send a random vector from a subspace  $\Gamma_i \subseteq \langle Y_i \rangle$ , where  $\Gamma_i$  is the space spanned by a random subset of the rows of  $Y_i$ . We denote the dimension of this subspace  $\dim(\Gamma_i)$  by *transmission degree*  $d$ , i.e.,  $d$  is the number of rows of the matrix  $Y_i$  that must be combined to form the outgoing packet. Clearly, the larger  $d$ , the higher the probability that this packet is not contained in the information

space of a neighbor and thus provides innovative information. Once the matrix  $Y_i$  has full rank, decoding can be done through Gaussian elimination. However, this is not necessary during the data dissemination process to form linear combinations. The main contribution of this paper is an analysis of the relationship among the network topology, the subspace overlap between any two given nodes, and the generation of innovative packets, as we now explain.

#### A. Innovative Information

Consider an all-to-all communication scenario, where each node has a packet to send and is interested in collecting the packets from all other nodes in the network. Let node  $i$  transmit a packet to node  $j$  from subspace  $\Gamma_i$  of dimension  $d$ . This packet is innovative if the vector contained therein and the rows of the matrix  $Y_j$  at the destination are linearly independent. We denote the dimension of node  $i$ 's space  $\dim(\langle Y_i \rangle)$  by  $\gamma_i$  and the dimension of the intersection of the spaces  $\dim(\langle Y_i \rangle \cap \langle Y_j \rangle)$  by  $\gamma_{i \cap j}$ . Given the transmission degree  $d = \dim(\Gamma_i)$  and the dimension of the intersection  $\gamma_{i \cap j}$ , we would like to characterize the probability that a random vector  $y_i$  is innovative for node  $j$ .

Let  $\psi_{i \rightarrow j}(d)$  be the probability that the packet  $y_i = m\Gamma_i$  generated from a  $d$ -dimensional subspace  $\Gamma_i$  of the information space of node  $i$  is innovative for node  $j$ , i.e.,

$$\psi_{i \rightarrow j}(d) = P(y_i \notin \langle Y_i \rangle \cap \langle Y_j \rangle).$$

$\psi_{i \rightarrow j}(d)$  can be computed in different manners, depending on the assumptions that we make on the content of the encoding vectors when the two nodes meet. In what follows, we discuss two possible models.

**Model A.** If we assume that the rows of the matrices  $Y_i$  and  $Y_j$  of the two nodes are elements picked uniformly at random from the whole  $N$ -dimensional space  $\mathbb{F}_q^N$ , we have:

$$\begin{aligned} \psi_{i \rightarrow j}(d) &= 1 - \frac{q^{\gamma_{i \cap j}} - 1}{q^{\gamma_i} - 1} \cdot \frac{q^{\gamma_{i \cap j} - 1} - 1}{q^{\gamma_i - 1} - 1} \cdots \frac{q^{\gamma_{i \cap j} - d + 1} - 1}{q^{\gamma_i - d + 1} - 1} \\ &= 1 - \frac{\begin{bmatrix} \gamma_{i \cap j} \\ d \end{bmatrix}_q}{\begin{bmatrix} \gamma_i \\ d \end{bmatrix}_q}, \end{aligned} \quad (1)$$

where  $\begin{bmatrix} N \\ d \end{bmatrix}_q$  is the Gaussian, or q-binomial, coefficient.

However, in practice, the elements of the matrices available at the nodes are not simply random elements of the complete information space. At the beginning of the dissemination process, the information to be disseminated is not mixed at all but only contains the nodes' own information. As the dissemination process proceeds, the information gets more and more mixed. This model does also not well capture the fact that nodes may perform Gaussian elimination during the dissemination process. In the extreme case where all information can always be decoded (and thus the rows in the coding matrices at the nodes contain all 'zeros' with a single 'one' at the corresponding column) we can reduce the buffer overlap problem to a classical balls and bins exercise.

**Model B.** In detail, we neglect that the matrices  $Y_i$  and  $Y_j$  (bins) may contain coded packets and assume all matrix rows (balls) are decoded.  $\gamma_{i \cap j}$  is thus the number of packets in common in the buffers of nodes  $i$  and  $j$ . The probability that  $i$  sends an innovative packet to  $j$  corresponds to the probability that  $i$  picks

a packet outside the set of common packets. Hence, it is easy to see that:

$$\psi_{i \rightarrow j}(d) = 1 - \frac{\begin{bmatrix} \gamma_{i \cap j} \\ d \end{bmatrix}}{\begin{bmatrix} \gamma_i \\ d \end{bmatrix}}, \quad (2)$$

where  $\begin{bmatrix} N \\ d \end{bmatrix}$  is the binomial coefficient.

**Innovative packet generation probability.** Using the total probability theorem, we can compute the probability that a transmitter node with a matrix of rank  $\gamma_i$ , linearly combining  $d$  packets from its own matrix, generates an innovative packet with respect to those available at the receiver. We call this probability the *innovative packet generation probability*, denoted by  $P_I$ . For both model A and model B, we have:

$$P_I = \sum_{\gamma=0}^N \mu(\gamma | \gamma_i, d) \psi_{i \rightarrow j}(d) \quad (3)$$

where  $N$  is the number of source packets (which equals the number of nodes in the network in our model), and  $\mu(\gamma | \gamma_i, d)$  is the probability that the size of the buffer overlap is  $\gamma$ , i.e.,  $\gamma_{i \cap j} = \gamma$ , for given  $\gamma_i$  and transmission degree  $d$ . In this paper  $\mu(\cdot)$  is obtained by simulation.

#### B. Buffer Overlap

Equations (1),(2) and (3) relate the buffer overlap to the innovative packet generation probability  $P_I$ . A quantity of interest, that is also needed for the calculation of  $P_I$  is the dimension of the overlap  $\gamma_{i \cap j}$ . In what follows, we characterize  $\gamma_{i \cap j}$  in terms of the fraction of the overlap between the matrices. First of all, we define  $m \triangleq \max\{0, \gamma_i + \gamma_j - N\}$  and  $M \triangleq \min\{\gamma_i, \gamma_j\}$ . Thus, we necessarily have:

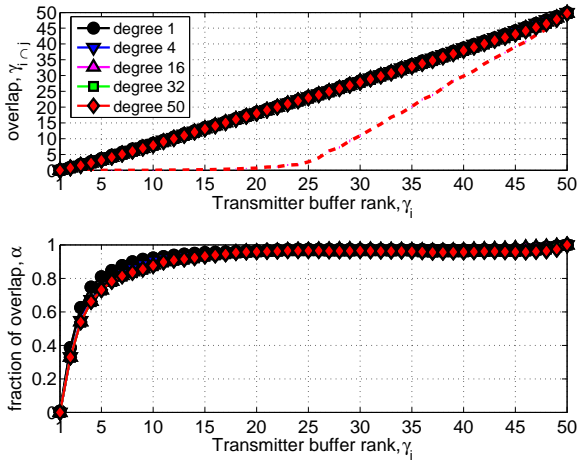
$$m \leq \gamma_{i \cap j} \leq M. \quad (4)$$

In fact, if  $\gamma_i + \gamma_j \leq N$ , the minimum value that  $\gamma_{i \cap j}$  can take is 0 (i.e., in case the subspaces generated by the buffers of nodes  $i$  and  $j$  do not overlap). Conversely, if  $\gamma_i + \gamma_j > N$  the minimum value that  $\gamma_{i \cap j}$  can take is  $\gamma_i + \gamma_j - N$  (i.e., the subspaces generated by the buffers of nodes  $i$  and  $j$  must overlap of at least this quantity). For  $M$ , the maximum value that  $\gamma_{i \cap j}$  can take is  $\gamma_i$  (resp.  $\gamma_j$ ) if the subspace generated by the buffer of node  $i$  (resp.  $j$ ) is a subset of the subspace generated by the buffer of node  $j$  (resp.  $i$ ). Thus, from (4)  $\gamma_{i \cap j}$  can be expressed as:

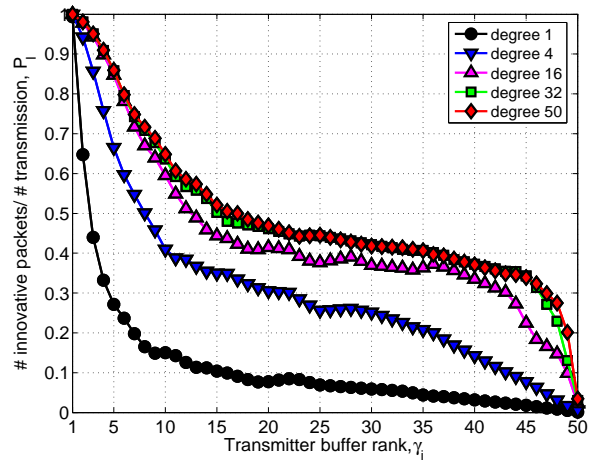
$$\gamma_{i \cap j} \triangleq m + \alpha(M - m), \quad (5)$$

where  $\alpha$  quantifies the relative overlap between the two buffers. It is worth noting that by definition  $0 \leq \alpha \leq 1$ . If  $\alpha = 0$  we have that  $\gamma_{i \cap j} = m$ , i.e., the overlap between the two buffers is the minimum possible. Conversely, if  $\alpha = 1$  it holds that  $\gamma_{i \cap j} = M$  and the overlap between the two buffers is maximized.

When  $M = m$ ,  $\alpha \triangleq 0$  if  $\gamma_i + \gamma_j \leq N$ , whereas if  $\gamma_i + \gamma_j > N$  we define  $\alpha \triangleq 1$ . Note that the previous definition of  $\alpha$  is consistent. From (5), when  $M = m$  the value of  $\alpha$  is undetermined. However, as  $\gamma_i$  is always greater than or equal to one, for  $M = m$  we can only have the following two cases: 1) when  $\gamma_i + \gamma_j \leq N$  from the definition of  $m$ , we get  $m = 0$ . If  $m = 0$ , from the definition of  $M$  we have that  $M = m$  if  $\gamma_j = 0$ , which implies  $\gamma_{i \cap j} = 0$ . Hence, defining  $\alpha = 0$  is correct as the two buffers have nothing in common in this case. 2) Along the same line,  $\gamma_i + \gamma_j > N$  implies  $m = \gamma_i + \gamma_j - N$ . As  $M = \min\{\gamma_j, \gamma_i\}$ , if  $M = \gamma_j$  from  $M = m$  we have  $\gamma_i = N$ ,

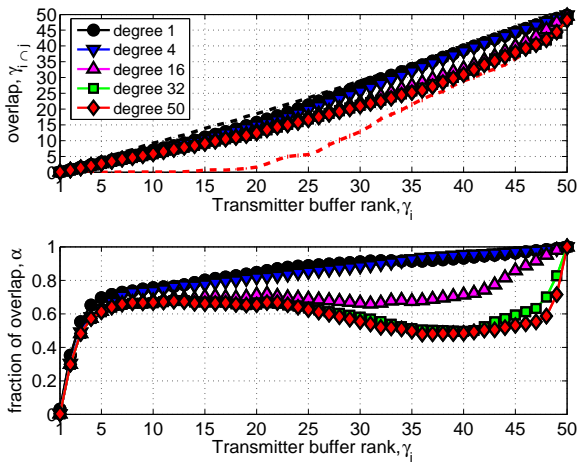


(a) Buffer overlap  $\gamma_{i \cap j}$  (top) and fraction of overlap  $\alpha$  (bottom)

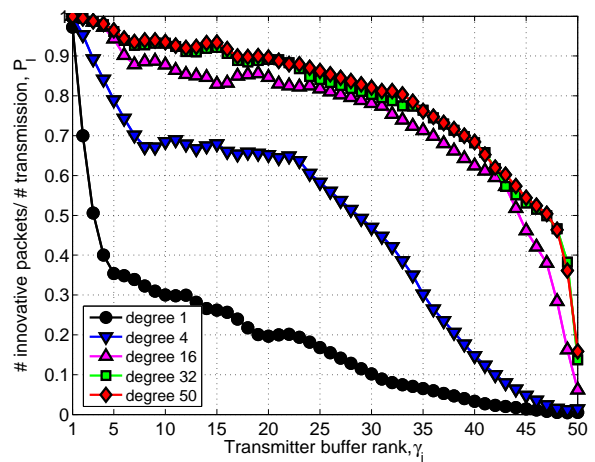


(b) Innovative Packet Generation Probability,  $P_I$

Fig. 1. Results for RNC in random static networks with  $N = 50$  nodes (scenario S1).



(a) Buffer overlap  $\gamma_{i \cap j}$  (top) and fraction of overlap  $\alpha$  (bottom)



(b) Innovative Packet Generation Probability,  $P_I$

Fig. 2. Results for RNC for a mobile network with low mobility and  $N = 50$  nodes (scenario S2).

otherwise if  $M = \gamma_i$  we have  $\gamma_j = N$ . Thus,  $\gamma_{i \cap j} = M = m$  and it is correct to define  $\alpha = 1$ . In section V, we characterize  $\alpha$  by simulation for four selected network scenarios.

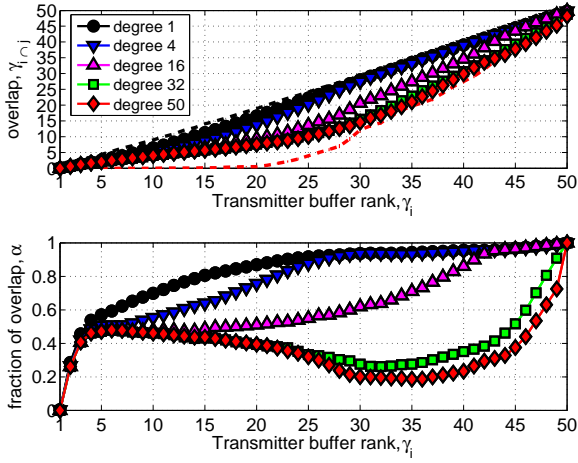
#### IV. SIMULATION DESCRIPTION

We consider four all-to-all communication scenarios: S1) a random static network, S2) a low mobility network, S3) a moderate mobility network, and S4) a mobile network with high mobility. Each of the  $N$  nodes has a packet to send and is interested in collecting the packets of all other nodes in the network. In the simulations, all nodes perform Gaussian elimination upon packet reception to ensure immediate decoding of the received information. We vary the dimension of the subspace from which packets are created (i.e., the transmission degree  $d$ ) from 1 to  $N$ . For the static scenario we consider topologies where  $N = 50$  nodes are placed uniformly at random (ensuring that full connectivity of the network is always guaranteed), with average node density of 8 neighbors per node. In low and moderate mobility scenarios, nodes additionally move according to a random waypoint mobility model with speeds uniformly distributed in the intervals  $[2, 4]$  and  $[8, 16]$  m/s (low and moderate mobility respectively). In the high mobility scenario (also called *random encounter mobility scenario*, see [2]), they move in a completely uncorrelated fashion and their speed is so high, as regard the network size, that the set of neighbors of any given

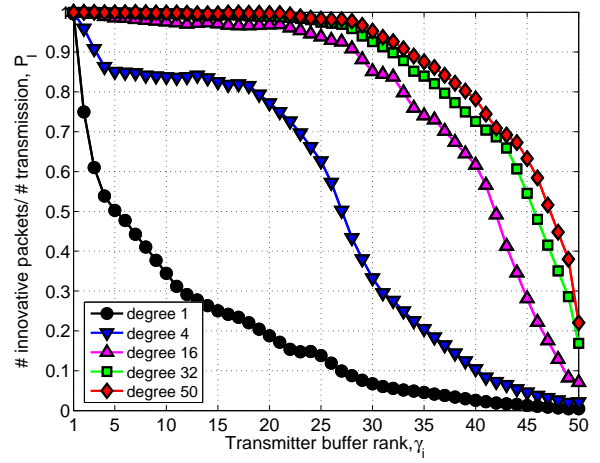
node at any time instant is independent from the neighbors of the same node at any other instant, i.e., the node's location is uncorrelated in time.

#### V. RESULTS AND COMMENTS

Figs. 1, 2, 3 and 4 show simulation results for random static networks (S1) and low (S2), moderate mobility (S3) and high mobility scenarios (S4), respectively. For the sake of readability, we plot only results for transmission degrees 1, 4, 16, 32 and 50. The plots on the left show the buffer overlap  $\gamma_{i \cap j}$  and the fraction of buffer overlap  $\alpha$ , whereas those on the right show the empirical  $P_I$ , that was measured by simulation as the ratio between the number of innovative packets delivered and the total number of transmissions (for given transmitter buffer rank and transmission degree). In Figs. 1(a), 2(a), 3(a) and 4(a) we also plot, by means of dotted lines, upper and lower bounds for  $\gamma_{i \cap j}$ , obtained averaging the values of  $M$  that were measured in the simulations with transmission degree 1 and averaging those of  $m$  with transmission degree 50 for the upper and the lower bound, respectively. It is worth observing that the gap between the two bounds narrows along the dissemination process (i.e., with increasing buffer rank). This is in line with the fact that, as packets are disseminated, the buffer rank increases for all nodes and, in turn, the probability of providing innovative information is reduced.

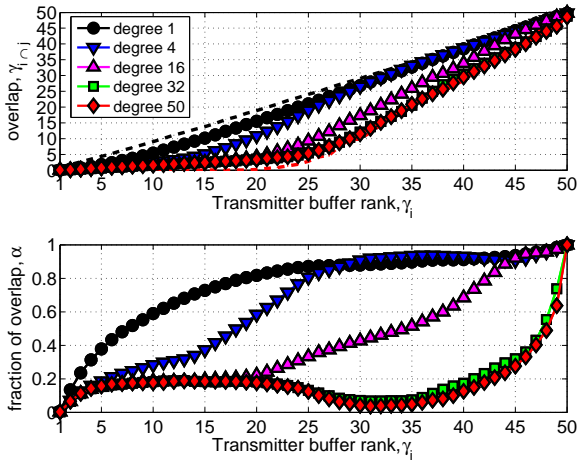


(a) Buffer overlap  $\gamma_{i \cap j}$  (top) and fraction of overlap  $\alpha$  (bottom)

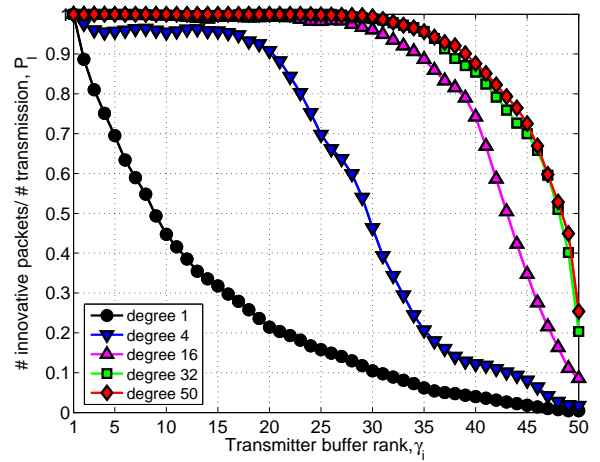


(b) Innovative Packet Generation Probability,  $P_I$

Fig. 3. Results for RNC for a mobile network with medium mobility and  $N = 50$  nodes (scenario S3).



(a) Buffer overlap  $\gamma_{i \cap j}$  (top) and and fraction of overlap  $\alpha$  (bottom)



(b) Innovative Packet Generation Probability,  $P_I$

Fig. 4. Results for RNC for a mobile network with high mobility with  $N = 50$  nodes (scenario S4).

For random static networks (Fig. 1) we see that the buffer overlap does not change, given a certain buffer rank of the transmitter, when using different transmission degrees (Fig.1(a)). Random networks where nodes communicate with the same few neighbors (i.e. sparse networks) make the diversity of information in the neighborhoods low. Therefore, a high transmission degree is necessary to achieve even a moderate probability of exchanging innovative packets, as shown in Fig. 1(b).

In Fig. 2 we show the same performance metrics for the low mobility scenario (S2). Here we get a first insight into the benefits of RLNC combined with mobility. The buffer overlap (Fig. 2(a)) for low transmission degrees gets close to the upper bound aforementioned, while for high transmission degrees and high transmitter buffer ranks it gets close to the lower bound. Compared to the static case, mobility lets nodes mix the information stored in their buffers and the diversity of information increases over the network as soon as new neighbors are encountered. Obviously, the lower the transmission degree, the higher the probability that a packet is generated from a subspace already available at some of the neighbors and is thus non-innovative. A low transmission degree also increases the probability that the packet can be decoded immediately in case it is innovative. As we have seen from the different probabilities of innovativeness associated with Model A and B, sending from a fully decoded

subspace may further decrease the probability of innovative information. The gain given by the low mobility in terms of  $P_I$  (Fig. 2(b)) is evident when compared to the random static case Fig. 1(b), due mainly to the mobility.

In Fig. 3 the higher mobility, compared to the low mobility scenario, increases the gap in terms of buffer overlap between low and high transmission degrees, as expected. Mobility helps particularly the high transmission degrees to get close to the maximum probability of sending innovative information, i.e. the benefits of full RLNC are well combined with moderate mobility (S3). This is particularly evident from the graph showing  $\alpha$ . At first,  $\alpha$  increases since information is mainly exchanged between neighbors. In the second half of the dissemination process,  $\alpha$  becomes very small as the overlap is close to the lower bound  $m$ .

In Fig. 4 we plot the results obtained in a random encounter scenario (S4). As expected, with high mobility the fraction of overlap  $\alpha$  decreases even further with respect to the other three cases, thus increasing  $P_I$ . We observe that both random coding and mobility help the mixing of information in the network through content diversification.

Finally, Fig. 5 shows the innovative packet generation probability  $P_I$  computed in conjunction with Model B (Eq. (2) and Eq. (3)) for the four scenarios. In all four cases, Model B

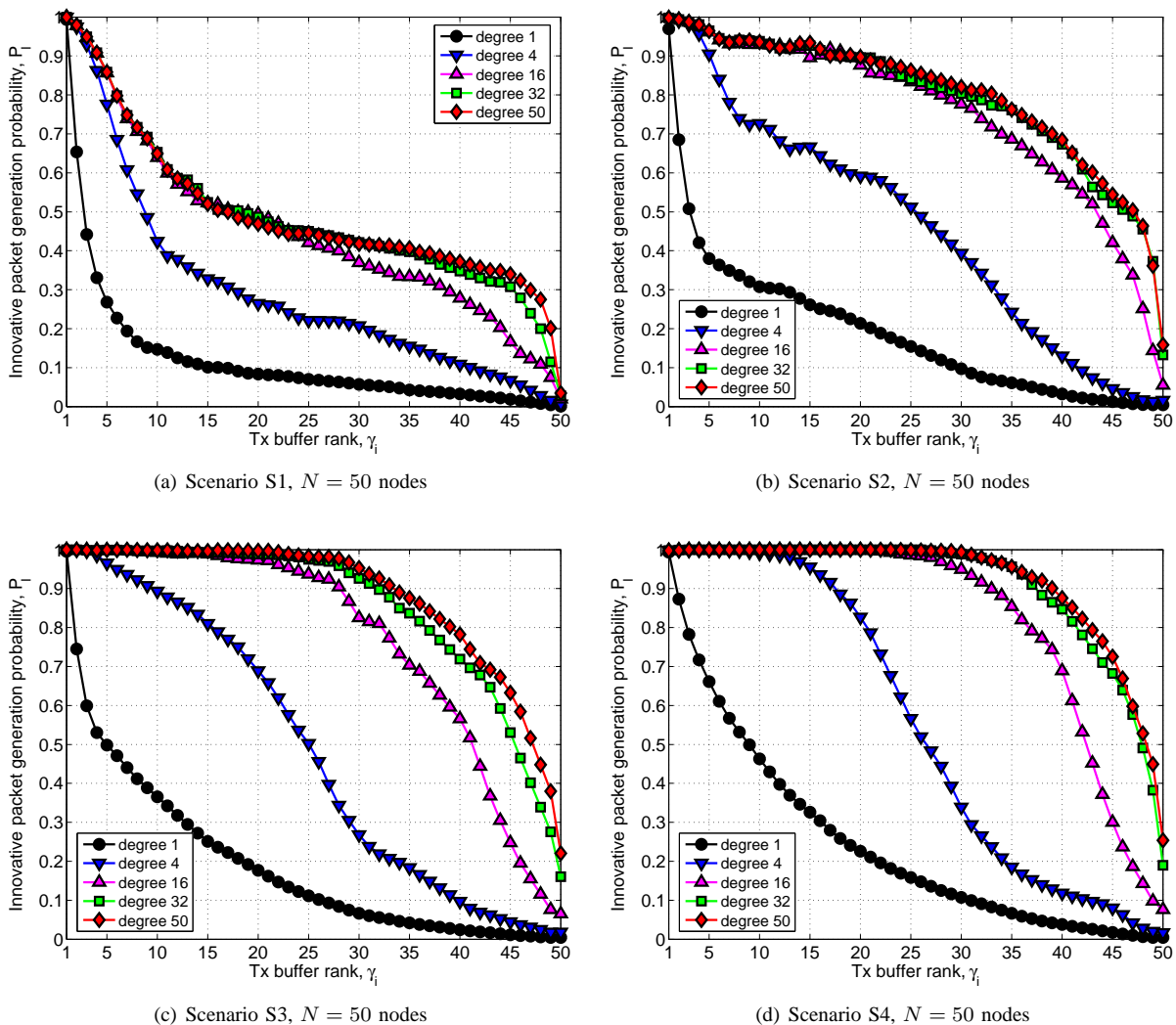


Fig. 5. Theoretical Innovative Packet Generation Probability,  $P_I$ .

captures extremely well what is observed in the actual simulations (compared to the plots showing  $P_I$  in Fig. 1, 2, 3 and 4). This confirms the validity of our model based on the probabilities of overlap  $\gamma_{i \cap j}$  and packet innovativeness  $\psi_{i \rightarrow j}(d)$ . The reasons for this good performance are twofold. As previously mentioned (section III), the information gets more and more mixed only as the dissemination process proceeds. In many cases, the amount of time in which the information is poorly mixed or not mixed at all is not negligible (with respect to the total dissemination time). Furthermore, nodes perform Gaussian elimination iteratively (i.e. after each received packet). Because of both of these aspects, the uniformity assumption that we discussed above in model A does not hold and the results of model B are closest to what we obtain in our simulations.

## VI. CONCLUSIONS

Given a network with nodes performing random linear network coding, this paper analyzed the relationship among the network topology, the subspace overlap between any two given nodes, and the generation of innovative packets. We proposed a framework to describe and understand this complex relationship. As the node mobility increases, the impact of transmission degree on the overlap between the subspaces increases as well. Node mobility helps the dissemination process in that packets are more likely to be innovative. Here, a transmission degree larger than 25% of the total number of source packets does not significantly add

to the dissemination performance. This behavior is well captured by the models we proposed in the paper. The analysis that we presented can be used in a more comprehensive framework with the objective of designing practical combination rules for RLNC in distributed and mobile networks. For future work, we intend to assess the validity of our model over a wider range of network sizes and topologies.

## REFERENCES

- [1] T. Ho, M. Médard, J. Shi, M. Effros, and D. R. Karger, "On Randomized Network Coding," in *41st Annual Allerton Conf. on Commun. Control and Computing*, Monticello, IL, US, Oct. 2003.
- [2] A. Kamra, V. Misra, J. Feldman, and D. Rubenstein, "Growth codes: Maximizing sensor network data persistence," in *ACM SIGCOMM*, Pisa, Italy, Sept. 2006.
- [3] M. Jafarisiavoshani, C. Fragouli, and S. Diggavi, "Subspace Properties of Randomized Network Coding," in *IEEE ITW*, Tahoe City, CA, US, Jul. 2007.
- [4] M. Jafarisiavoshani, C. Fragouli, S. Diggavi, and C. Gkantsidis, "Bottleneck Discovery and Overlay Management in Network Coded Peer-to-Peer Systems," in *IEEE INM*, Kyoto, Japan, Aug. 2007.
- [5] S. Acedanski, S. Deb, M. Médard, and R. Koetter, "How good is random linear coding based distributed networked storage?" in *NetCod*, Riva del Garda, Italy, Apr. 2005.
- [6] J. Liu, Z. Liu, D. Towsley, and C. Xia, "Maximizing the data utility of a data archiving & querying system through joint coding and scheduling," in *IPSN 2007*, Cambridge, MA, US, Apr. 2007.
- [7] D. Munaretto, J. Widmer, M. Rossi, and M. Zorzi, "Resilient Coding Algorithms for Sensor Network Data Persistence," in *IEEE EWSN*, Bologna, Italy, Jan./Feb. 2008.
- [8] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network Coding," *IEEE Trans. on Information Theory*, vol. 49, no. 2, Feb. 2003.