# Inter-Session Network Coding in Delay Tolerant Mobile Social Networks: an Empirical Study

Neetya Shrestha and Lucile Sassatelli
Université Nice Sophia Antipolis, Laboratory I3S, CNRS UMR 7271
06903 Sophia Antipolis, France
Email: {shrestha,sassatelli}@i3s.unice.fr

***Abstract*—Delay Tolerant Mobile Social Networks (DTMSNs) are networks made of human-carried wireless devices with intermittent connections, and whose physical meeting patterns make cluster into social communities. In such environments, routing is a challenge as the limited resources (like memory and contact opportunities) must be efficiently used and shared between the sessions (or users, contents). To handle several unicast sessions, Inter-Session Network Coding (ISNC) has been proven necessary for optimal throughput in general networks, but is a delicate problem as it can quickly get detrimental. This paper investigates empirically whether ISNC can be beneficial in DTMSNs. We first show that on a simple chain topology, without or with a hub node, no gain can be generally obtained when contacts are bidirectional. We then show that if non-directionality impedes ISNC gain, it can be due to greedy replication, and on the same DTMSN operated with a social-aware routing algorithm, the set of chosen routes turns ISNC into beneficial. On a butterfly topology, we investigate the impact, on ISNC gain, of key network parameters such as buffer management, copy (memory) budget and network load. This allows to determine what parameters to take into account when designing a decentralized ISNC criterion for general topologies.**

## I. INTRODUCTION

Delay (or disruption) Tolerant Networks (DTNs) are networks made of wireless nodes with intermittent connections. When the devices are carried by humans, the mobility exhibits heterogeneous patterns where node clustering into communities arises owing to social relationships [1]. We refer to these DTNs as Delay Tolerant Mobile Social Networks (DTMSNs). The three main goals of DTMSNs in civilian applications can be deemed as: (i) to provide network access to remote communities (e.g., Bytewalla [2]), (ii) provide cheaper content access by file exchange in ad hoc mode (e.g., PSN [3], [4], Liberouter [5]), (iii) to offload the cellular networks (e.g., [6], rescue operations).

A number of opportunistic utility-based routing algorithms (coping with long disconnections) that leverage the diversity of social ties have been proposed, e.g., [7], [8]. To improve the probability of delivery within a certain deadline, several copies of the same packet can be disseminated, and this benefit may further improve with Network Coding (NC) which has attracted an increasing interest for DTNs [9]. NC is a networking paradigm that is a generalization of routing [10], aiming in particular at improving throughput and resiliency to topology changes. There are two versions: (i) intra-session NC, mixing only the packets of the same session, (ii) inter-session NC (ISNC) mixing packets possibly pertaining to different sessions. ISNC is necessary to achieve optimal throughput in general (see [11] and references therein), but is a delicate problem because to retrieve its intended packets, a destination node needs to receive also other sessions' packets, called 'remedy packets' thereafter. If it does not, then ISNC can degrade performance as compared to routing. Whether ISNC can bring some gain in DTMSNs with unicast sessions, and how, is the subject of this article. The optimization problem of ISNC for multiple unicast sessions has been proven NP-hard [10]. A number of works (e.g., [11], [12]) have come up with approximate solutions for static directed networks. When coming to DTNs, there is a priori no reason for considering that two nodes can exchange packets in a single direction. For undirected networks, Li and Li in [13] have shown theoretically that, for the multicast problem, the throughput increase ratio between intra-session NC and half-integer routing is upper-bounded by two, and even less with fractional routing. These results for multicast and intra-session NC are readily transposable to ISNC with unicast sessions. The non-directionality of DTNs is hence a first hurdle to the possible gain with ISNC. However, ISNC has proven very attractive in undirected wireless mesh networks [14], specifically owing to the time-shared wireless medium. So one can think that the constrained shared resources (buffer, contact opportunities) in DTNs can make ISNC attractive too, despite the non-directivity. But the second difficulty to readily apply this reasoning to DTNs is that there is no radio interference owing to the low node density and radio range.

The goal of this paper is therefore to identify what can be the advantages of ISNC in DTMSNs, and how they are impacted by key network parameters such as routing, Buffer Management (BM), copy budget and network load. We take an empirical approach, considering simple topologies of node communities. Our contributions are: (i) We first show that on a simple chain topology, without or with a hub node, no gain can generally be obtained by ISNC when contacts are bidirectional. In particular, we quantify the impact of the inter-community meeting intensity on the ISNC gain. (ii) We find that if non-directionality impedes ISNC gain, it can be due to greedy replication, and on the same DTMSN operated with a social-aware routing algorithm, the set of chosen routes turns ISNC into beneficial. This is shown by building on the SimBet routing strategy [7] and opportunistically using as remedy packets the copies wandering in other communities than their destination's. (iii) On a butterfly topology, we investigate the impact, on ISNC gain, of the copy (memory) budget and network load. We thereby determine what parameters to take into account when designing a decentralized ISNC criterion, and how.

**Related works**: In [15], NC is considered at some intermediate hub nodes, but only across packets destined to the same destination node. In [16], Zhang *et al.*, consider both intra- and inter-session NC in homogeneous DTNs. For unicast sessions with different sources and destinations, uncontrolled ISNC is shown not to perform better than intra-session. Here, we tackle the more general problem of ISNC for unicast sessions with different destinations. In [17], we have presented a pairwise ISNC policy and an analytical model where the number of packets per session can be any, corresponding to the case where a file is split into several packets, and the metric (whether delay or delivery probability) is on the whole file. Here, we address the optimization problem expressed in [17] by reducing the parameter space to identify sound heuristics that will serve in the final goal of designing fully decentralized social-aware ISNC policies, which is a future work.

## II. NETWORK MODEL AND NOTATION

In order to investigate the possible advantages of ISNC in DTMSNs, we consider the multi-community model of [8]. The network is made of $N$ nodes clustered into $C$ communities, each made of $N_c$ nodes, $c \in \{1, C\}$. The meetings of any pair of nodes are assumed to be Poisson distributed, and the inter-meeting intensity is defined as the mean number of meetings per time unit. It is assumed to be the same, $\beta_{ii}$, for all pairs of nodes pertaining to community $i$, while $\beta_{ij}$ denotes the inter-meeting intensity of any pair of nodes in community $i$ and $j$. The concept of community imposes that $\beta_{ii} > \beta_{ij}$, for all $i \neq j$. The graph nodes do not represent network nodes anymore, but entire node communities. We consider the network bears $R$ unicast sessions with source-destination node pairs $(S_i, D_i)$, $i \in \{1, R\}$, and $P_i(\tau)$ is the probability that $D_i$ has obtained its information packets by time $\tau$. A session is made of $K$ packets. Let $U(.)$ be any classical utility concave function, taken as $\log(1 + x)$ here. If $R = 2$, then we defined the utility over both sessions as $obj(\tau) = U(P_1(\tau)) + U(P_2(\tau))$. The nodes' buffer size is denoted by $B$ (in packets). A single packet can be sent in each direction upon each meeting. The maximum number of copies of a packet index (whose payload keeps the same or gets modified under NC [9]) that can be in the network at any instant of time is denoted by $M$. In case session $i$ is made of a single packet, then the latter is denoted by $P^{(i)}$.

## III. THE CHAIN TOPOLOGY

### A. A chain without a hub node

We first consider the simplest toy topology depicted in Figure 1.a, made of only 2 communities, with $R = 2$ reverse sessions. The rationale for considering this topology is that the shared constrained resource is the relays' buffers. So ISNC may help serve faster both destinations if ISNC can be incorporated into the BM policy. Indeed, in the destination community, the BM without ISNC can be such that the packets not destined to this community are erased by those which are. However, doing so might prevent the packets originating in this community to spread enough so as to reach their destination community. In such a case, ISNC can be thought to bring some gain. Therefore, a possible improvement by ISNC is closely tied to the BM policy.

Let us hence first analyze the BM problem without ISNC. A number of works have analyzed the BM (and scheduling for buffer sizes greater that 1) in DTMSNs, such as [18], [19]. In particular, [18] derives an optimal drop policy that discriminates packets based especially on their number of copies and TTL. Below, we chose not to implement this refined BM, thereby leaving more odds to ISNC for outperforming a non ISNC policy. We consider Spray-and-Wait replication, $K = 20$, $B = 1$, $\beta_{11} = \beta_{22} = 0.05$, $N_1 = N_2 = 50$, and $p$ is the probability that a packet of session $i$ replaces a packet of session $j \neq i$ in community $C_{D_i}$, for $i, j \in \{1, 2\}$. Figure 2 shows the probability that each destination has received at least 10 (different) information packets by time $500s$, when the meeting intensity $\beta_{12}$ between both communities varies. This quantity is obtained by the fluid model described in [17], and matches well the simulations of this simple setting. We observe that the optimal value of the overwrite probability $p$ varies from 0.3 to 0.1 when $\beta_{12}$ decreases. Indeed, the lower $\beta_{12}$, the slower the propagation of packets between communities. So once a packet reaches its destination community $C_D$, it must not propagate too fast so as to give time to all the information packets generated and hold by $C_D$ to cross to their destination community, before getting erased. In order to maximize an objective defined as a sum of concave functions of the delivery probabilities of each session, the best is indeed to equalize them, i.e., to equally share the common bandwidth, as depicted in Figure 1.b. One can then wonder whether ISNC can help share the bandwidth between both reverse sessions in a more efficient way, that is, if the BM was given an extra choice that is instead of replacing packets, mixing them thereby generating ISNC packets, would that bring any gain in utility (by serving both communities with the same packet)?

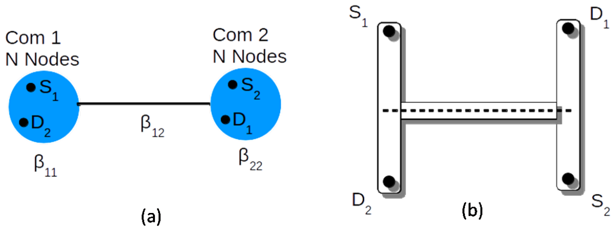


Fig. 1. (a) A 2-community network. (b) Bandwidth sharing with 2 reverse sessions.

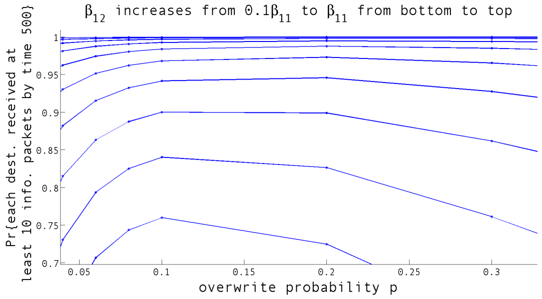


Fig. 2. Impact of $\beta_{12}$ on the optimal overwrite probability $p$.

In order to investigate this question, we consider $B = 1$, $K = 1$ and no packet drop leaving more odds to ISNC to be beneficial.

*Lemma 3.1:* Let $Y_2$ denote the number of nodes holding $P^{(2)}$ in community 2 when no ISNC is employed, and $\alpha =$

$\beta_{12}/\beta_{22}$. A necessary condition for ISNC to outperform no ISNC is:

$$\frac{Y_2}{N_2} > \frac{3+\alpha}{4} \ . \qquad (1)$$

*Proof*: See Appendix A. ◇

When the communities 1 and 2 are merged, that is $\alpha = 1$, no gain can be expected and we get back the result of Zhang *et al.* in [16]. There might exist values of $\alpha < 1$ that lead to a gain of ISNC. In practice, we have not been able to find such values of $\alpha$, for 1 and more packets per session. In particular, the above lemma may hold for several packets per session if we replace the definition of $Y_2$ by the total number of session 2 packets with intra-session NC, and similarly for the other quantities involved in the proof of the lemma. Having identified that ISNC is likely not to bring some gain in such a 2-community topology, we next consider a topology closer to COPE [14].

### B. A chain with a hub node

As aforementioned, a well-known application of ISNC is for ad hoc wireless mesh networks [14] (Figure 3.a). ISNC can help save transmissions even though they are bi-directional, owing to the constrained resource sharing (the wireless medium is shared in time). Thus, we consider a similar community-based topology depicted in Figure 3.b, as a simplistic DTMSN where ISNC might be beneficial. The nodes apart across the two communities can only exchange through their (asynchronous) meetings with the hub node.

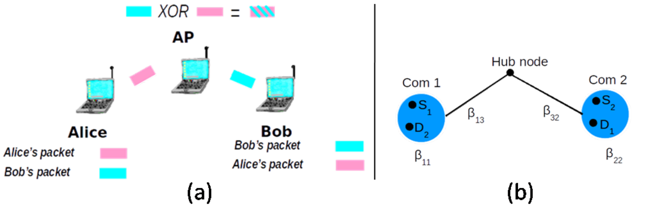


Fig. 3. (a) The COPE example [14]. (b) A chain topology with a hub node.

The rationale for considering this topology as a good candidate to enable ISNC gain is the following: (i) if the hub node's buffer size is high enough, then there is no competition for buffer access between both sessions, and ISNC is not expected to bring any advantage; (ii) if the hub node's buffer size is constrained, say equal to one packet in the extreme case, then it may be beneficial for both sessions to allow the hub node to carry a coded packet that may serve both sessions, depending on what community the hub node meets next.

In order to investigate the possible gain, we carry out simulations on this topology, with $K = 1$, $B = 1$ and the following BM: (i) when no relay node is allowed to mix both session packets, each time the hub node gets a copy of any packet, it overwrites that it may have held, and $P^{(i)}$ overwrites in $C_{D_i}$ all the copies of $P^{(j)}$, $i \neq j$; (ii) when the hub node holds $P^{(i)}$ and meets with $P^{(j)}$, then it stores the XoR of both packets. This coded packet overwrites in $C_{D_i}$ the copies of $P^{(j)}$, but not those of $P^{(i)}$.

We consider two cases: (i) in the unidirectional case, when two nodes meet, only one of them can transmit to the other one; (ii) in the bidirectional case, both meeting nodes can transmit to each other. Figure 4 shows the utility value obtained in the bidirectional and unidirectional case without and with ISNC. All the simulation results in this paper are averaged over 30 runs and the 5% confidence intervals are plotted. We observe that ISNC brings some gain only in the unidirectional case. This is explained as follows. In the bidirectional case, if the hub node holds $P^{(i)}$ and meets with a node in $C_{D_i}$ that holds $P^{(j)}$, then it delivers (overwrites) $P^{(i)}$ and gets $P^{(j)}$. Therefore, ISNC is not useful (its impact within one of the side community can be analyzed as in Section III-B). In this topology, ISNC can bring some gain only if one transmission of the XoRed packet may serve both sessions, after it got generated, that is possible only in the unidirectional case. The same holds if several relay hubs are assembled within a daisy chain. So we need to move on to another topology to make ISNC gain appear in general bidirectional transmissions. As the transmissions are not simultaneous in two different communities, as mentioned above, ISNC saving is in terms of occupied buffer space, so we need a transfer in the same buffer in the same direction for both sessions. This implies that $S_1$ and $D_2$ nodes cannot be in the same community anymore.

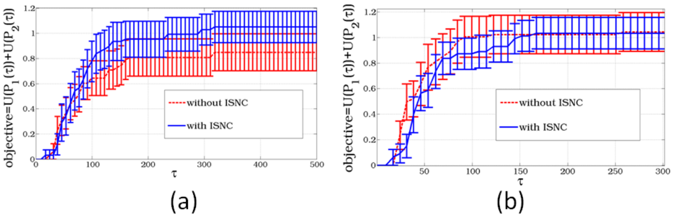


Fig. 4. Objective value with and without ISNC, $K = 1$, $M = 200$, (a) unidirectional, (b) bidirectional

## IV. THE BUTTERFLY TOPOLOGY

### A. Impact of routing on ISNC gain

To remedy the impediments of the above kind of topology in allowing ISNC gains, we consider the well-known butterfly topology depicted in Figure 5 with the two sessions with source-destination pairs as $(S_1, D_1)$ and $(S_2, D_2)$. We run experiments, with $M = 200$ and Spray-and-Wait (i.e., greedy, the copies are spread to the first met nodes epidemically). In order to favor ISNC, we take the inter-meeting intensities as depicted in Figure 5, with $x = 5.10^{-3}$. Their ratios are such that the link between the hub node and the destination communities is the bottleneck, while it is easy for the destinations to get the remedy packets thanks to the high inter-meeting intensities on the side links.

The numerical results, we do not include here, show no improvement in utility $obj(\tau)$. The reason is as follows. The equivalent static graph of communities is undirected and epidemic routing yields the packet routes depicted in Figure 5.a. So ISNC cannot help even with a constrained buffer size (of 1 packet) at the hub node because the main paths taken by both sessions are opposite, and we get back to the situation described in the last section. This is in line with the results of Li and Li in [13] who showed that the throughput increase brought by NC in undirected networks vanishes in front of fractional routing. ISNC would be useful if the routes depicted in Figure 5.b were taken.

It hence turns out that the possible advantage of ISNC in this synthetic DTMSN, whose communities form a butterfly

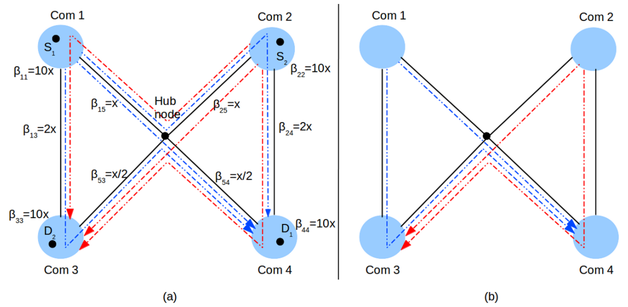


Fig. 5. (a) All possible routes are depicted in blue and red for each session 1 and 2, respectively. (b) The set of single routes allowing to get a gain with ISNC.

graph but where contacts are inherently bidirectional and asynchronous, is closely tied to the routes the packets flow through. However, the way of operating such DTMSNs is usually not to flood the packets to as many as possible nodes, but rather to focus the limited amount of copies to certain well-chosen relay nodes. Thus, routing in DTMSNs shall not allow to take all the possible routes.

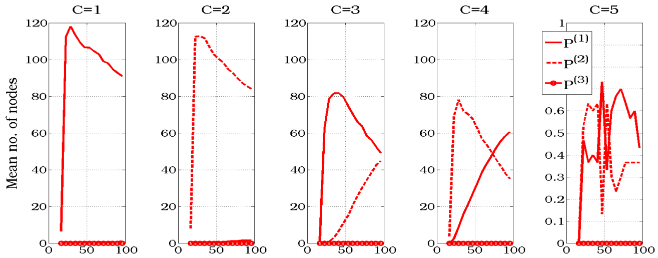
We verify this assumption by analyzing the routes taken by the packets when routing is performed with SimBet [7]. SimBet is a utility-based routing algorithm, that flows packets through relays with high utility towards the destination. When multiple copies can spread, a counter is kept in the packet copy header, and the copy budget shared upon replication is based on the meeting nodes' utilities. Figure 6.a depicts the evolution over time of the number of each packet type in each community. Sessions and topology are symmetric, so we comment hereafter only for session 1. Packet $P^{(1)}$ (solid line) first spreads inside its source community $C_{S_1} = 1$, then reaches $C_{D_1} = 4$ mostly through the hub node as we see that the increase in $P^{(1)}$-infected nodes in $c = 4$ precedes the increase in $c = 2$, while the hub node gets readily infected. Community $c = 2$ remains almost uninfected by $P^{(1)}$. This shows that the routes taken by $P^{(1)}$ and governed by SimBet are very close to those in Figure 5.b, identified as the routes susceptible to benefit from ISNC. This motivates our approach of using ISNC in DTMSN operated with a social-aware routing algorithm. Below we show that on the set of routes selected by the SimBet routing algorithm, ISNC can be beneficial.

## *B. Combining ISNC and social routing*

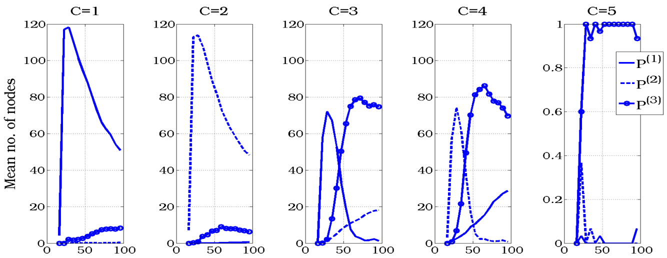
We now study how to enable ISNC when the routes are chosen by a social routing algorithm (such as SimBet [7]). To do so, two distinct but correlated problems arise:

- **P1** Where (and when) must the coding be performed?
- **P2** How to make the destinations get the remedy packets they need?

Problem P1 relates to the choice of what nodes are allowed to mix messages of what sessions. These nodes can be either set fixed, and this gets us a proof of concept of ISNC gains we analyze thoroughly in this paper on the butterfly topology with SimBet routing, or decided online based on local decisions at nodes, and this is the very next step, out of the scope of this paper. P1 is related to P2 because, depending on if and when the remedy packets can reach the destination nodes, the nodes allowed to send mixed packets can differ.



(a) without ISNC



(b) with ISNC (at hub node)

Fig. 6. Mean number of nodes infected with each type of packets in each community.

Problem P2 can be solved in two different ways: (i) either a signaling mechanism is set up, such as that of Eryilmaz and Lun [11] or Kreishah *et al.* [12], to explicitly inform the source nodes to send their packets to destinations they are not destined to, (ii) or no signaling is performed and only the remedy packets that are inherent overhead produced by the social routing algorithm are opportunistically exploited to allow ISNC. Let us note that, doing so, not all coding opportunities can be leveraged, but there is a trade-off between the additional complexity and the ISNC gain. In this work, we investigate only the latter approach (ii) for the sake of simplicity.

## *C. Buffer and Copy Counter Management*

We set that the only node allowed to mix packets is the hub node, and it does so as soon as it can, i.e., a XoRed packet named $P^{(3)}$ can be generated at node $H$ only if: (i) $H$ already holds $P^{(1)}$, and has no more room in its buffer, and (ii) $H$ meets $A$ holding $P^{(2)}$ (or the other way around), and (iii) SimBet triggers the transmission of $P^{(2)}$ to $H$, based on utilities only, and (iv) the BM below allows the replacement of $P^{(1)}$ by $P^{(3)}$ at $H$. The copy counter assigned to $P^{(3)}$ is then the sum of the counter of the replaced packet and the copy budget handed over by $A$, determined by the copy share of SimBet [7]. We consider a BM which cannot favor ISNC compared with single routing and bias the results (the symmetric holds for session 2):

**F1** Destination node $D_1$ can erase $P^{(1)}$ from all nodes (in any community) but cannot erase $P^{(2)}$. A packet $x$ (IS or nonIS) can erase another packet $y$ in a community $c$ only if $x$ is destined to $c$ and $y$ is neither destined to $c$ nor has its source in $c$. For ISNC, $P^{(1)}$ can replace $P^{(3)}$ in $C_{D_1}$.

**F2** Keep on spreading the energy budget even though the payload of the already-there packets does not change. For example, when node $A$ with $P^{(1)}$ meets $B$ with $P^{(3)}$, if SimBet utility would trigger transmission of a copy budget to $B$, then it gets added to the counter at $B$, although the payload remains the same. This feature, allowed by the use of ISNC, allows to re-focus the copy budget through coded packets to better serve both sessions. When the spray counter of a copy drops to zero, it is dropped.

**F3** Destination node $D_1$: (i) erases $P^{(3)}$ upon reception from a node in $C_{D_1}$, and (ii) signals to the nodes of outside communities that it has received $P^{(3)}$ and/or recovered $P^{(1)}$.

Let us specify that the above items constitute BM choices aimed at fairly comparing ISNC with non-ISNC policies. We can also consider the scheme proposed by Krifa et al. in [18] whose tractability, for DTMSNs, is based on the assumption that the node inter-meeting intensities stem from the same distribution, and the copies are spread over numerous enough different nodes. For the sake of ease of interpretation, and because it is likely that such assumptions do not hold for the highly heterogeneous topologies considered here, we did not implement it here. For other topologies, we can envision such a policy to replace F1, and hence apply both to no ISNC and ISNC policies.

We consider the topology and sessions of Figure 5.a, still with $K = 1$, with $N_c = 250$ nodes, for $c = 1, \dots, 4$, community 5 is the single hub node, $\beta_{13} = \beta_{24} = x$, $\beta_{15} = \beta_{25} = 10x$, $\beta_{53} = \beta_{54} = 8x$ and $\beta_{cc} = 15x$, for $c = 1, \dots, 4$, with $x = 5.10^{-4}$. We use 15% of the simulation duration as warm-up phase to provide an opportunity to gather information about the nodes within the network, as in [7]. After the warm-up phase, the messages are allowed to disseminate in the network. We plot in Figure 7 the objective function $obj(\tau)$ defined in Section II. In order to assess the impact of each component F1, F2 and F3, the results incrementally adding each of those are shown. The curve label "woIS" refers to the case where SimBet routing alone is used, without any session mixing. In this case, only F1 can apply. We point out that here intra-session NC is not mentioned as messages are made of a single packet. We show that with such a systematic coding at the hub node in the butterfly topology operated with SimBet, ISNC outperforms single routing. The remaining 3 curves (labeled "wIS") show that without specific additional buffer or copy counter management, ISNC bring some gain, incrementally improved over F1, F2 and F3. The ISNC gain is in particular explained by Figure 6.b, where we can see that the hub gets occupied by a coded packet ($P^{(3)}$) systematically. This hub node is indeed the point of congestion, as it is at crossroads and has buffer size of 1.

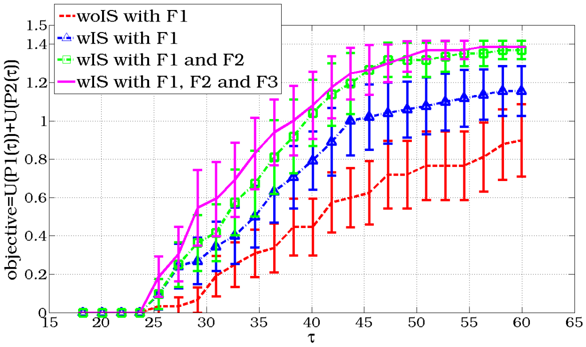
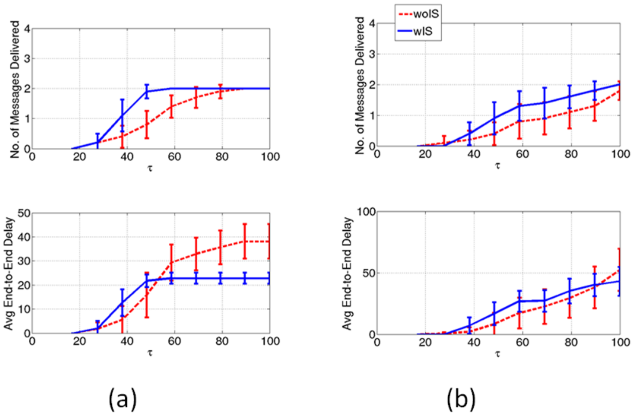


Fig. 7. Objective value for the different buffer and copy counter management (2 crossed sessions).

### D. Impact of the copy budget

We now analyze how the ISNC gain varies with the energy/memory constraint, that is with the maximum number of copies per packet $M$. Figure 8 depicts the performance in terms of two metrics, of two crossed sessions, with $M = 200$ and $M = 40$, respectively. Without ISNC, the message delivery delay is the time for the intended destination to get the first copy. With ISNC, the message delay is the time for the destination node to recover its intended message, either by receiving the original (uncoded) packet, or by receiving a coded packet and its remedy packet (then the delay is the maximum of the time when both coded and remedy packets are received). We observe in the upper-left Figure 8 the performance corresponding to that of Figure 7. For lower $M$, the gain of ISNC decreases, owing to the lower probability for the destination node to get the remedy packet. Hence, when designing a decentralized coding criterion in future work, we will need to pay specific attention to the constrained copy budget to decide whether to trigger session mixing.



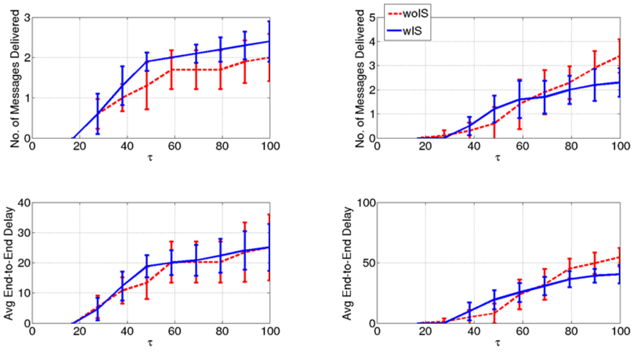
(a) (b)

Fig. 8. $R = 2$, crossed sessions, (a) $M = 200$, (b) $M = 40$

### E. Impact of the number of concurrent sessions

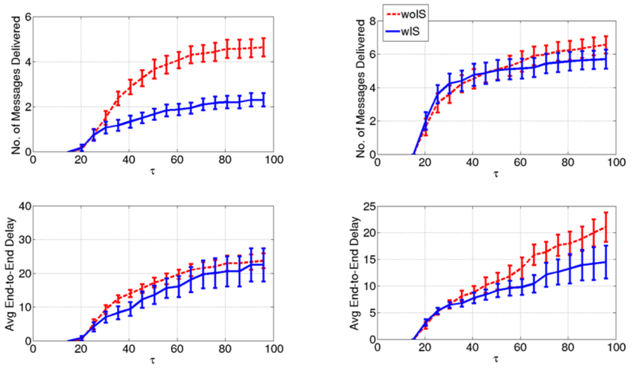
We now analyze the impact of the network load on ISNC gain, where the network load is considered as the number of concurrent sessions. Let us specify that we only allow pairwise ISNC, meaning that a coded packet mixes at most two sessions. When the number of sessions increases, to prevent the network from being stuck, we add two features to the BM described earlier: (i) if the buffer is full, the packet can be overwritten by the incoming one with a certain probability that depends on the counters of each packet, and (ii) each time a packet (either replica or coded) is created at a node, it is assigned an exponentially distributed TTL whose mean decreases with its copy counter. The details are provided in Appendix B.

Figures 9 and 10.a show the relative performance of ISNC at the hub node for $R = 4$ and $R = 10$ crossed sessions respectively, that is half of the sessions has S-D pairs in the communities $(1, 4)$ while the remaining half has in $(2, 3)$. We observe that, for a given $M$, the ISNC gain decreases as $R$ increases, and turns into a detriment for $R = 10$. We verify that this is due to the remedy packets that do not make it early enough at the destination because of the higher contention for buffer access. Figure 10.b shows that the degradation brought by ISNC dwindles when the sessions' S-D pairs are picked uniformly at random over the communities. While the number of delivered messages is not higher with ISNC, the delivery delay is lower than no ISNC.



(a) (b)

Fig. 9. $R = 4$, crossed sessions, (a) $M = 200$, (b) $M = 40$



(a) (b)

Fig. 10. $M = 200$, (a) $R = 10$, crossed sessions, (b) $R = 10$, homogeneous sessions.

A difficulty of devising efficient ISNC schemes is that ISNC must perform at least as well as routing (possibly with intra-session NC [9]), and not degrade the performance. The results above show that both the copy budget and the network load are important parameters that must be taken into account when designing a decentralized social-aware ISNC scheme.Let us mention that a metric we did not investigate here is node fairness. Indeed, how the nodes that are key hubs in the networks are loaded with others' traffic is an important concern for the sustainability of DTMSNs. A follow-up work hence consists in comparing ISNC with SimBet policy [7] and FOG [20] in trading off fairness and efficiency.

## V. CONCLUSION

We have shown experimentally, on synthetic contact traces corresponding to simple topologies, that ISNC can be beneficial in DTMSNs, and under what conditions. The gains are specifically permitted when a subset of possible routes are selected by a utility-based algorithm. The reach of this work extends beyond these simple topologies as it has unveiled possible cases (in terms of topology and network parameters) where ISNC, coupled with social routing, is beneficial. The next challenge is to detect online when these cases arise in real-world traces, that is designing a decentralized coding criterion allowing each node that is presented with a coding opportunity to decide whether to code or not, based only on the local information it has gathered. Devising such an online ISNC policy is the topic of our ongoing work.

## REFERENCES

[1] T. Hossmann, T. Spyropoulos, and F. Legendre, "A complex network analysis of human mobility," in *IEEE NetSciCom 11 (co-located with IEEE INFOCOM 2011)*, Shanghai, China, Apr. 2011, pp. 876 – 881.

[2] "Bytewalla DTN on Android Phones." [Online]. Available: http://sourceforge.net/projects/bytewalla/

[3] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *ACM SIGCOMM workshop on DTN (WDTN)*, Aug. 2005, pp. 244–251.

[4] E. Yoneki, P. Hui, S. Chan, and J. Crowcroft, "A socio-aware overlay for publish/subscribe communication in delay tolerant networks," in *ACM MSWiM*, Chania, Crete Island, Greece, Oct. 2007.

[5] "Liberouter: Towards autonomous neighborhood networking." [Online]. Available: http://www.ict-scampi.eu/results/scampi-liberouter/

[6] L. Keller, A. Le, B. Cici, H. Seferoglu, C. Fragouli, and A. Markopoulou, "Microcast: Cooperative video streaming on smartphones," in *ACM MobiSys*, Low Wood Bay, UK, Jun. 2012.

[7] E. M. Daly and M. Haahr, "Social network analysis for information flow in disconnected delay-tolerant MANET," *IEEE Trans. on Mob. Comp.*, vol. 8, no. 5, pp. 606 – 621, May 2009.

[8] T. Spyropoulos, T. Turletti, and K. Obraczka, "Routing in delay-tolerant networks comprising heterogeneous node populations," *IEEE Trans. on Mob. Comp.*, vol. 8, no. 8, pp. 1132 – 1147, Aug. 2009.

[9] Y. Lin, B. Li, and B. Liang, "Efficient network coded data transmissions in disruption tolerant networks," in *IEEE Int. Conf. on Comp. Comm. (INFOCOM)*, Phoenix, USA, Apr. 2008, pp. 1508–1516.

[10] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Trans. on Netw.*, vol. 11, no. 5, pp. 782–795, 2003.

[11] A. Eryilmaz and D. S. Lun, "Control for inter-session network coding," in *Inf. Th. and Appl. workshop (ITA)*, San Diego, CA, Jan. 2007.

[12] A. Khreishah, W. Chih-Chun, and N. B. Shroff, "Rate control with pairwise intersession network coding," *IEEE/ACM Trans. on Netw.*, vol. 18, no. 3, pp. 816–829, Jun. 2010.

[13] S. Maheshwar, Z. Li, and B. Li, "Bounding the coding advantage of combination network coding in undirected networks," *IEEE Trans. on Info. Th.*, vol. 58, no. 2, pp. 570– 584, 2012.

[14] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORs in the air: Practical wireless network coding," *IEEE/ACM Trans. on Netw.*, vol. 6, no. 3, pp. 497 – 510, 2008.

[15] S. Ahmed and S. S. Kanhere, "Hubcode: hub-based forwarding using network coding in delay tolerant networks," *Wireless Comm. and Mob. Comp.*, vol. 13, no. 9, p. 828846, May 2011.

[16] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Benefits of network coding for unicast applications in disruption-tolerant networks," *IEEE/ACM Trans. on Netw.*, vol. 21, no. 5, pp. 1407–1420, Oct. 2013.

[17] N. Shrestha and L. Sassatelli, "On control of inter-session network coding in delay-tolerant mobile social networks," in *ACM MSWiM*, Montreal, QC, Canada, Sep. 2014, pp. 231–239.

[18] A. Krifa, C. Barakat, and T. Spyropoulos, "Message drop and scheduling in DTNs: Theory and practice," *IEEE Trans. on Mob. Comp.*, vol. 11, no. 9, pp. 1470 – 1483, Sep. 2012.

[19] A. Balasubramanian, B. Levine, and A. Venkataramani, "DTN routing as a resource allocation problem," *ACM SIGCOMM Comp. Comm. Rev.*, vol. 37, no. 4, pp. 373–384, Oct. 2007.

[20] A. Mtibaa and K. A. Harras, "FOG: Fairness in mobile opportunistic networking," in *IEEE Conf. on Sensor, Mesh and Ad Hoc Comm. and Netw. (SECON)*, Salt Lake City, UT, 2011, pp. 260–268.

# APPENDIX

## A. *Proof of Lemma 1*

Without IS-NC:

$$Pr\{S_1 \text{ be useful to } D_1 \text{ within } \Delta t | \text{a node holds } S_1\}$$

$$.Pr\{\text{a node holds } S_1\} = \beta_{22}\Delta t(1-\frac{Y_2}{N}),$$

$$Pr\{S_1 \text{ be useful to } D_2 \text{ within } \Delta t | \text{a node holds } S_1\}$$

$$.Pr\{\text{a node holds } S_1\} = 0,$$

With IS-NC:

$$Pr\{S_3 \text{ be useful to } D_1 \text{ within } \Delta t | \text{a node holds } S_3\}$$

$$.Pr\{\text{a node holds } S_3\} \leq \beta_{22}\Delta t\frac{Y_2}{N}(1-\frac{Y_2'}{N}),$$

$$Pr\{S_3 \text{ be useful to } D_2 \text{ within } \Delta t | \text{a node holds } S_3\}$$

$$.Pr\{\text{a node holds } S_3\} \leq \beta_{12}\Delta t\frac{X_1}{N}(1-\frac{Y_2'}{N}),$$

where, assuming $\Delta t$ low enough:

- $\beta_{22}\Delta t$ represents the probability that this coded packet hits $D_1$ within $\Delta t$ from the time it has been generated in community-2, either by a combination of $S_1$ and $S_2$ or by replication.
- $\frac{Y_2}{N}$ (resp. $\frac{X_1}{N}$) represents the probability that $D_1$ (resp. $D_2$) has received $S_2$ (resp. $S_1$). This is an upper-bound on the actual probability as, assuming that the number of community-2 nodes infected (that have met) with session-2 packets is $\frac{Y_2}{N}$ at the end of $\Delta t$, $\frac{Y_2}{N}$ is the probability that a community-2 node holds at least one packet of session-2, not necessarily $S_2$. Packets are indeed sent continuously by both sessions.
- $Y_2$ (resp. $Y_2'$) denotes the steady-state fraction of nodes holding $S_2$ in community-2 when no IS-NC is employed (resp. when it is), and $Pr\{\text{a node holds } S_1\} = X_2/N = (1 - Y_2/N)$ is the fraction of community 2 nodes holding $S_1$ in the steady-state (it is optimal for the success probabilities to occupy all the network nodes in the steady state). With IS-NC however, community-2 nodes can carry either $S_1$, $S_2$ or $S_3$, therefore $(1 - Y_2'/N)$ is an upper-bound for $Pr\{\text{a node holds } S_3\}$.

Then,

$$Pr\{S_1 \text{ be useful to } D_1 \text{ or } D_2 \text{ within } \Delta t\} \leq \beta_{22}\Delta t(1-\frac{Y_2}{N}),$$

$$Pr\{S_3 \text{ be useful to } D_1 \text{ or } D_2 \text{ within } \Delta t\} \leq \beta_{22}\Delta t\frac{Y_2}{N}\dots$$

$$+\beta_{12}\Delta t\frac{X_1}{N} - \beta_{12}\beta_{22}\frac{Y_2}{N}\frac{X_1}{N}(1-\frac{Y_2'}{N})(\Delta t)^2.$$

Assuming that the last term is neglected as each component factor is by assumption lower than 1, we can translate the condition

$$Pr\{S_3 \text{ be useful to } D_1 \text{ or } D_2 \text{ within } \Delta t\}$$

$$> Pr\{S_1 \text{ be useful to } D_1 \text{ or } D_2 \text{ within } \Delta t\}, \tag{2}$$

into the necessary condition:

$$(\beta_{22}\Delta t\frac{Y_2}{N} + \beta_{12}\Delta t\frac{X_1}{N})(1-\frac{Y_2'}{N}) > \beta_{22}\Delta t(1-\frac{Y_2}{N}),$$

that is

$$-(\frac{Y_2'}{N})^2 + \frac{Y_2'}{N} > \frac{1}{1+\alpha}(1-\frac{Y_2}{N}),$$

where, $\alpha = \beta_{12}/\beta_{22}$. Owing to the fact that $-x^2 + x \leq 0.25$ for all $x \in \mathbb{R}$, a necessary condition to satisfy the above inequality is:

$$\frac{Y_2}{N} > \frac{3+\alpha}{4} \tag{3}$$

◇

## B. *Buffer management for several sessions*

**F4** Choice between packet $A$ and $B$ to be kept when two nodes $I$, with packet $A$, and $J$ with packet $B$ meet and have full buffers: if $A$ (resp. $B$) is in its destination community and $B$ is not (resp. $A$), then we do not take any action, otherwise we pick up a random variable $X$ such that

$$Pr(X=0) = \frac{\alpha * n_A}{n_A + n_B},$$
$$Pr(X=1) = \frac{\alpha * n_B}{n_A + n_B},$$
$$Pr(X=2) = 1-\alpha.$$

where $n_A$ and $n_B$ are the copy counters of packets $A$ and $B$ respectively, $\alpha$ is constant such that $\alpha : [0,1]$. Then, we take following action based on different values of $X$:

$X = 0$ : $A$ overwrites $B$, while $B$ cannot overwrite $A$,

$X = 1$ : $B$ overwrites $A$, while $A$ cannot overwrite $B$,

$X = 2$ : there is no exchange of packets and no overwrite is performed.

Furthermore, each message has a Time-To-Live (TTL) value, after which the message is dropped to allow new messages that arrive at a node to occupy the buffer space. If $L_n >= T\_start + V$ , drop the packet $P_1$ from node's buffer, where $T\_start$ is the time when a copy of packet $P_1$ is received, $V$ is TTL value and $L_n$ is the current time when a decision is to be taken if the packet is to be dropped or not. We can select the value of $V$ in following 3 different ways:

1) Constant: TTL is considered as a constant value; $V = V_0$.
2) Exponential: TTL is exponentially distributed with some constant mean, $meanV$ such that $V = exprnd(meanV)$ where $meanV = V_0$.
3) Exponential and dependent on copy counter: TTL is exponentially distributed with some mean value that is dependent on the number of copies yet to spread, $current\_counter$ such that $meanV = \frac{V_0 * current\_counter}{M}$.