On Control of Inter-session Network Coding in Delay-Tolerant Mobile Social Networks

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 (or disruption) Tolerant Networks (DTNs) are networks made of wireless nodes with intermittent connections. In such networks, various opportunistic routing algorithms have been devised so as to cope with the lack of contemporaneous end-to-end route between a source and a destination. We consider DTNs made of mobile nodes clustered into social communities, with unicast sessions. Network coding is a generalization of routing that has been shown to bring a number of advantages in various communication settings. In particular, inter-session network coding (IS-NC) is known as a difficult optimization problem in general. In this article, we introduce a parameterized pairwise IS-NC control policy for heterogeneous DTNs, that encompasses both routing and coding controls with an energy constraint. We derive its performance modeling thanks to a mean-field approximation leading to a fluid model of the dissemination process, and validate the model with numerical experiments. We discuss the optimization problem of IS-NC control in social DTNs. By showing numerical gains, we illustrate the relevance of our approach that consists in designing IS-NC control policies not reasoning on specific nodes but instead on the coarse-grained underlying community structure of the social network.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Store and forward networks*

General Terms

Algorithms, Measurement, Performance

Keywords

Delay Tolerant Networks; Mobile Opportunistic Networks; Routing Algorithms; Network coding; Control policy

1. INTRODUCTION

Delay (or disruption) Tolerant Networks (DTN) are sparse Mobile Ad-hoc NETworks (MANETs) that can rely neither on any infrastructure to support communication nor on the guarantee that a path exists between a source and a destination at any instant of time. Sparsity arises owing to short radio range, obstruction or intermittent sleeping mode.

In order to achieve end-to-end communication in the absence of a contemporaneous path between the source and the destination at the time of the transmission, the nodes rely on the store-carry-and-forward paradigm, taking advantage of the transmission opportunities between relays. The packet delay can be lowered by spreading multiple copies of the same packet. The possible ways to spread multiple copies have been investigated in several proposals [34, 31]. In particular, Spray-and-Wait (SaW) [31] has been proposed to achieve a trade-off between network resource (memory and energy) consumption and performance.

In this paper we focus on mobile social DTN, that are Pocket Switched Networks (PSN) formed by people carrying portable devices [14, 9]. Target applications include vehicular ad hoc networks [26], Publish-Subscribe [5] and content-centric systems for DTNs [36]. Human mobility exhibits heterogeneous patterns where node clustering into communities arises owing to social relationships [13]. A characteristic of DTNs is the distribution of the node inter-contact durations. In this article, a DTN with identically (resp. non-identically) distributed inter-contact times across the nodes is referred to as "homogeneous" (resp. "heterogeneous") DTN. Social DTNs are hence heterogeneous DTNs.

To improve the benefit of disseminating redundant packets, coded redundant packets can be generated by the relays instead of or additionally to replicated copies, that is performing Network Coding (NC) in DTN. NC is a networking paradigm that is a generalization of routing [1, 18]. Specifically, random NC [12] has attracted an increasing interest for DTNs [35, 22]. The benefits are increase in throughput, as well as adaptability to network topology changes. There are two types of NC: in intra-session NC, only the packets belonging to the same session are coded together (i.e., combined), while in inter-session NC (IS-NC) packets pertaining to different sessions can be combined. IS-NC is necessary to achieve optimal throughput in general (see [8] and references therein) but represents a difficult optimization problem, in particular for DTNs, as detailed in Section 2.

Contributions:

- We design a parameterized pairwise IS-NC control policy for heterogeneous DTNs, that encompasses both routing and coding controls with an energy constraint. We present the resulting dissemination protocol.
- We derive its performance modeling thanks to a mean-field approximation leading to a fluid model of the dissemination process. We validate the model by numerical experiments.

MSWiM'14, September 21–26, 2014, Montreal, QC, Canada.

http://dx.doi.org/10.1145/2641798.2641833.

- We discuss the optimization of IS-NC control policy benefits in social DTNs, and show that the fluid model can be used to devise such optimal policy that jointly exploits the nodes' social acquaintances and the IS-NC. By showing numerical gains, we illustrate the relevance of our IS-NC control policy that is based on the coarse-grained community structure rather on individual nodes.

This paper does not aim at presenting a self-contained decentralized IS-NC protocol that can be confronted with existing routing policies in DTN. It aims at devising, modeling and proving the benefit of a centralized social-aware (community-based) pairwise IS-NC policy. Hence, the next step after this work is to study numerically the optimization problem in order to extract heuristics to devise a decentralized IS-NC policy for social DTNs.

This article is organized as follows. Section II presents related works. Section III presents the network model and the IS-NC protocol. In Section IV, we derive the performance modeling, and we discuss the optimization problem in Section V, then illustrate the relevance of controlling IS-NC based on communities to obtain some gains. Section VI concludes the paper discussing limitations and future works.

2. RELATED WORK

For homogeneous DTN, several works have considered intra-session NC which is now well understood in this case. Lin et al. in [22] investigated the use of intra-session NC using the SaW algorithm and analyzed the performance in terms of the bandwidth of contacts, the energy constraint and the buffer size. However, neither background traffic nor other running session are assumed beside the unicast session of interest. In [28], we have lifted this assumption and modeled information dissemination of several concurrent unicast sessions in homogeneous DTNs, when IS-NC and SaW routing are employed. In the present article, we extend this work not only to heterogeneous DTNs to predict the performance of contending unicast sessions, either inter-session network coded or not, but also model the control of IS-NC decisions based on the social features of the DTNs.

On the other hand, a number of routing policies have been proposed for heterogeneous DTNs to improve the trade-off between performance and resource consumption. Their principle is not to spend the allowed number of transmissions with the first met nodes, contrary to SaW [31], but instead to smartly choose the relays to give the copies to, in terms of the network social features. We can cite BubbleRap [15] and SimBet [6], where some global and local ranks are used for each node to orientate and control the spreading. In [32], Spyropoulos et al. introduced the mobility model we consider, and also used a fluid model to prove that one of the utility-based replication policies they consider achieves lower delivery delay than greedy. The optimality of such forwarding policies based on such a model is investigated in [29].

In [2], NC is considered at some intermediate hub nodes, but only across packets destined to the same destination node. In [37], Zhang *et al.* consider both intra- and inter-session NC in homogeneous DTNs. For unicast sessions with different sources and destinations, uncontrolled IS-NC is shown not to perform better than intra-session. In this paper we tackle the more general problem of control of pairwise IS-NC for unicast sessions with different destinations. When considering several unicast sessions, IS-NC can bring throughput and fairness gains [33, 17] both on lossless and lossy links. However, the optimization problem of IS-NC for multiple unicast sessions has been proven NP-hard [18], in particular because of the joint problems of subgraph selection and coding decisions, that can be solved independently for a single multicast session [23]. Therefore, all the works addressing the problem of IS-NC target suboptimal, yet continually improved, methods [16, 33, 8, 17]. These approaches are not directly applicable to DTNs as they assume fixed topologies and may incur heavy signaling. In particular, in [8], Eryilmaz and Lun introduced a routing-scheduling-coding strategy using back-pressure techniques. Modeling the coded flows as "poisoned" [33], the queue-length exchange is meant to determine the location of the encoding, decoding, and remedy generating nodes. The control policy of IS-NC and its modeling we introduce here, allow to account simultaneously for coding and remedy packets, that is the fact that a node may send to a destination packets not destined to it so as to help decoding mixed packets. Furthermore, all these works considered directed networks. However, there is a priori no reason for considering that two nodes can exchange packets in a single direction in DTNs. Li and Li in [20] have shown theoretically what can be the maximum throughput improvement with intra-session NC in undirected networks, compared with what can be obtained with integral, half-integer and fractional routing. In particular, for the multicast problem, the throughput increase ratio is upper-bounded by two between NC and half-integer routing, or even less with fractional routing [20, 24]. However, the shared resources (buffer, contact bandwidth) in DTNs make IS-NC attractive as in wireless mesh networks [16], though these networks are undirected. Hence, we are tackling the open problem of IS-NC design in social DTN, and want to study, thanks to the tunable pairwise IS-NC control policy and its performance model introduced in the present paper, what improvement can be brought by IS-NC, and how.

3. NETWORK MODEL

We consider the heterogeneous mobility model introduced in [32, 4]: the network is made of N mobile nodes divided into C communities such that $N = \sum_{i=1}^{C} N_i$ where N_i is the number of nodes in community i, and we assume that a node pertains to only one community. Table 1 gathers the main parameters' notation used throughout the paper. In this model, the time between two consecutive contacts is exponentially distributed with a certain mean. The accuracy of this model has been discussed in [11] and shown for a number of mobility models (Random Walker, Random Direction, Random Waypoint). The inter-meeting intensity β_{ij} is defined as the inverse of this mean and represents the mean number of contacts per time unit between a given node of community i and another given node of community j. We assume that $\beta_{ii} > \beta_{ij}$, for $i \neq j$, for all $i, j \in \{1, \ldots, C\}$. The matrix β storing the $\{\beta_{ij}\}_{i,j=1}^{C}$ defines the inter-meeting intensity of any pair of nodes.

Two sources S_1 and S_2 of communities s_1 and s_2, respectively, want to send a file each to their respective destinations D_1 and D_2 in communities d_1 and d_2, respectively. We assume that the file to be transferred needs to be split into K packets: this occurs owing to the finite duration of contacts among mobile nodes or when the file is large with respect to the buffering capabilities of the nodes. The message is considered to be well received if and only if all the K packets of the source are recovered at the destination. We do not assume any feedback.

We assume that the bandwidth, defined as in [22] as the number of packets that can be exchanged during a contact in each direction (thereby accounting both for the rate and the contact duration), is stochastic and follows any known distribution of mean Bw. The buffer size is assumed to be any known integer, denoted by B, equal for all the nodes in the network. Note that these assumptions are not necessary for the dissemination protocol presented in Algo. 1-2 to work.

Symbol	*Meaning*
Network settings	
N	total number of nodes excluding the sources and the destinations
C	number of node communities
N_i	number of nodes in community i
β_{ij}	inter-meeting intensity of a node in community i with a node in community j
Bw	bandwidth: mean number of packets that can be exchanged during a contact in each direction
Communication settings	
S_1, S_2	source node of session 1, 2
D_1, D_2	destination node of session 1, 2
K_1, K_2	number of information packets of session 1, 2
K_1', K_2'	maximum number of packets that can be released by S_1, S_2
M, Q	maximum number of copies of an index released by S_1, S_2
S_{11}, S_{22}	set of indices associated to pure payloads sent out by source S_1, S_2
S_{31}, S_{32}	set of indices emitted by S_1 (resp. S_2) associated to a mixed payload, that a combination from pure payloads of S_1 and S_2
X_{ic}, Y_{ic}	number of nodes in community c that carry i indices in S_{11} (resp. S_{22})
Z_{ic}^1, Z_{ic}^2	number of nodes in community c that carry i indices in S_{31} (resp. S_{32})
$\tilde{X}_{Ic}, \tilde{Y}_{Ic}$	number of nodes in community c that carry index I of S_{11} (resp. S_{22})
$\tilde{Z}_{Ic}^1, \tilde{Z}_{Ic}^2$	number of nodes in community c that carry index I of S_{31} (resp. S_{32})
$u_{ce}^{11}(t)$, $u_{ce}^{22}(t)$	probability that a node of community c gives a packet with an index in S_{11} (resp. S_{22}) to a node of community e upon meeting at time t, provided that it is possible to copy such an index (it exists at the sending node and its spray-counter is below M (resp. Q))
$u_{ce}^{31}(t)$, $u_{ce}^{32}(t)$	probability that a node of community c gives a packet with an index in S_{31} (resp. S_{32}) to a node of community e upon meeting at time t, provided that it is possible to copy such an index (it exists at the sending node and its spray-counter is below M (resp. Q))
l	$l = \sum_{i=11,22,31,32} l_i$ for a $(c, \mathbf{l})$-node

Table 1: Main notation used throughout the paper

3.1 Inter-session NC

Let us now describe simply how a node having two packets of two different sessions, performs IS-NC to forge a new coded packet to be sent out. The process described hereafter is depicted in Fig. 1. All nodes can identify the session number of each packet. Consider that packets P_1 and P_2, belonging to sessions S_1 and S_2, are Random Linear Combinations (RLC) of the K_1 and K_2 original information packets, respectively. The header coefficients of P_1 and P_2 are hence K_1-long and K_2-long, while payloads are L_1 and L_2-long, where L_1 and L_2 are the maximum size of packets of S_1 and S_2, respectively. The packet resulting from an RLC of P_1 and P_2 has header coefficients $(K_1 + K_2)$-long, and payload $\max(L_1, L_2)$-long. The original K_1 packets of session 1 can be recovered if and only if the matrix made of the coding coefficients can undergo a Gauss-Jordan elimination resulting in only elements of the K_1-size identity matrix over the K_1 columns assigned to session 1 and for the corresponding rows, all the other columns are zero. Thereafter, the number of received Degree of Freedom (DoF) of session 1 is the number of identity elements over these K_1 columns.

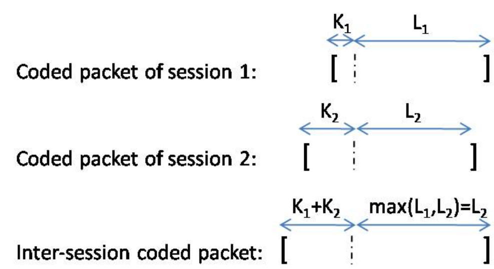

Figure 1: Generation of an inter-session network coded packet

3.2 The parameterized pairwise IS-NC control policy

We build on the buffer structure for intra-session NC framework employed with SaW [21], that we modify to account for IS-NC of two sessions. Note that binary SaW is considered for implementation, but any other spraying with a token mechanism allowing to control the dissemination, such as that of [37], can be considered for implementation and modeling. At a relay node buffer, a packet is associated with 3 fields: index, spray-counter, payload. Both sources S_1 and S_2 spread $K_1' \geq K_1$ (resp. $K_2' \geq K_2$) RLCs over the K_1 (resp. K_2) information packets. Each RLC sent out by S_1 (resp. S_2) is associated with an index in $S_{11} = \{1, \ldots, K_1'\}$ (resp. $S_{22} = \{K_1' + 1, \ldots, K_1' + K_2'\}$) and with a counter M (resp. Q). A node is said to be a $(c, \mathbf{l})$-node, if it belongs to community c and has in its buffer $\mathbf{l} = (l_{11}, l_{22}, l_{31}, l_{32})$ indices of S_{11}, S_{22}, S_{31} and S_{32}, respectively. Also, we take $l = \sum_i l_i$. In order to control the IS-NC decisions, we introduce the policy $\mathbf{u}_{ab}(t) = (u_{ab}^{11}(t), u_{ab}^{31}(t), u_{ab}^{22}(t), u_{ab}^{32}(t))$, for all a and b in $\{1, \ldots, C\}$. It corresponds to the probability to draw, at each transmission opportunity, an index of each kind for sending it from a node of community a to a node of community b. As aforementioned, the number of transmission opportunities at each contact has an arbitrary distribution with mean Bw. The components of $\mathbf{u}_{ab}$ are defined in Table 1. We constrain $\sum_i u_{ab}^i \leq 1$ to allow for a node not to spread as many packets as possible if it is better to keep some transmissions for later meetings. Algo. 1-2 describes the dissemination protocol we consider with IS-NC. The cases where the sources or destinations are met can be found in Algo. 1.

DEFINITION 3.1. *An index I is said to be part of S_{11} or S_{31} (resp. S_{22} or S_{32}) if either I or $I-(K_1'+K_2')$ are in $[1, K_1']$ (resp. I or $I-(K_1'+K_2')$ are in $[K_1'+1, K_1'+K_2']$).*

For this protocol, we consider that no packet exchange is possible once the buffer of the receiving node is full, but note that it is straightforward to adapt the framework to any kind of buffer management policy (including possible exchange when the buffers are full). As well, D_1 (resp. D_2) is allowed to make a meeting relay drop only packets of S_{11} (resp. S_{22}) it already got. It is worth noting also that the binary split of the counter in between the replicated packet and that remaining in the sending node is not a constraint of the framework, and can be changed for any other counter sharing, such as that based on some utility and presented in [6].

4. MODELING DISSEMINATION

The goal is to predict the evolution over time of the different numbers of nodes describing the dissemination process and defined in Table 1. To do so, we resort to a mean-field approximation that allows to predict the mean behavior of a system, modeled as a Markov chain, made of a growing number of interacting objects. Owing to the lack of space, we do not provide a formal proof that such a mean-field approximation hold in the present case, but rather only a sketch of the proof, and assess numerically the accuracy of the model in Section 4.3. By Theorem 3.1 of [19], the quantities $X_{ic}, Y_{ic}, Z^1_{ic}, Z^2_{ic}, \tilde{X}_{Ic}, \tilde{Y}_{Ic}, \tilde{Z}^1_{Ic}, \tilde{Z}^2_{Ic}$ defined in Table 1, that are random processes depending on the random mobility process, can be approximated by deterministic processes that are the solutions of certain coupled Ordinary Differential Equations (ODEs). These ODEs stem from the limit of the system dynamics (the "drift") for large N and are called the fluid model. Let us now present the ODEs of the fluid model.

Below we present the main components of the model and in particular we highlight how the control of IS-NC over the time and the communities is taken into account. As in the protocol description above, the notation in what remains corresponds to considering a node A that may send packets to a node B that is it meeting.

4.1 Evolution of the buffer occupancy distribution

The ODEs for X_{ic} and Z^1_{ic} write as:

$$dX_{ic}(t) = \beta_{s_1c}N_c \sum_{\mathbf{l}^B} P_j(c,\mathbf{l}^B)P_{gs11}(i-l^B_{11},\mathbf{l}^B,\mathbf{u}_{s_1c})\dots$$

$$+\beta_{cd_1}N_c \sum_{\mathbf{l}^B} P_j(c,\mathbf{l}^B)P_{ls11}(l^B_{11}-i,\mathbf{l}^B,\mathbf{l}^{D_1},\mathbf{u}_{cd_1})\dots$$

$$+N_c\sum_{e=1}^{C}\sum_{\mathbf{l}^A,\mathbf{l}^B}\beta_{ec}N_eP_j(c,\mathbf{l}^A)P_j(c,\mathbf{l}^B)\dots$$

$$P_{grs11}(i-l^B_{11},\mathbf{l}^A,\mathbf{l}^B,\mathbf{u}_{ec})\dots$$

$$-\beta_{cs_1}N_c \sum_{\mathbf{l}^B:l^B_{11}=i,p_{11}>0} P_j(c,\mathbf{l}^B)P_{gs11}(p_{11},\mathbf{l}^B,\mathbf{u}_{s_1c})\dots$$

$$-\beta_{cd_1}N_c \sum_{\mathbf{l}^B:l^B_{11}=i,p_{11}>0} P_j(c,\mathbf{l}^B)P_{ls11}(p_{11},\mathbf{l}^B,\mathbf{l}^{D_1},\mathbf{u}_{cd_1})\dots$$

$$-N_c\sum_{e=1}^{C}\sum_{\substack{p_{11}\mathbf{l}^A,\\ \mathbf{l}^B:l^B_{11}=i}}\beta_{ec}N_eP_j(c,\mathbf{l}^A)P_j(c,\mathbf{l}^B)\dots$$

Algorithm 1: Protocol with IS-NC - Part 1

Data: a $(a,\mathbf{l}^A)$-node A (i.e., in community a with $\mathbf{l}^A$), and a node $(b,\mathbf{l}^B)$-node B, $\mathbf{u}_{ab}=(u^{11}_{ab},u^{22}_{ab},u^{31}_{ab},u^{32}_{ab})$, the number of packet transmission opportunities w
Result: How many and what packets generated by A to be stored at B

Let $Hd(\mathbf{u})$ be the distribution of a discrete random variable Y with 4 values, value i being taken with probability u^i_{ab}.
if $A == S_1$ *and* $B \neq D_1$ *and* $B \neq D_2$ **then**
    Draw q_{11} from a binomial distribution (w, u^{11}_{ab}).
    if *Num_of_RLCs_sent_by_*$S_1 < K_1'$ **then**
        Send $x = \min(q_{11}, K_1' -$ *Num_of_sent_RLCs_by_A*$, B - l^B)$ RLCs with indices from *Num_of_sent_RLCs_by_*$S_1 + 1$ up to *Num_of_sent_RLCs_by_*$S_1 + x$
else if $A == S_2$ *and* $B \neq D_1$ *and* $B \neq D_2$ **then**
    Draw q_{22} from a binomial distribution (w, u^{22}_{ab}).
    if *Num_of_RLCs_sent_by_*$S_2 < K_2'$ **then**
        Send $x = \min(q_{22}, K_2' -$ *Num_of_sent_RLCs_by_A*$, B - l^B)$ RLCs with indices from *Num_of_sent_RLCs_by_*$S_2 + 1$ up to *Num_of_sent_RLCs_by_*$S_2 + x$
else if $B == D_1$ **then**
    Drop the packets with indices of S_{11} already present at D_1. Update l^A_{11} accordingly.
    $x = 0$;
    while $x \leq w$ **do**
        Draw y from $Hd(\mathbf{u})$.
        if y==*11* **then**
            Send 1 packet whose index is in S_{11} to D_1 and drop it from A.
        if y==*22* **then**
            Send 1 packet whose index is in S_{22} to D_1.
        if y==*31* **then**
            Send 1 packet whose index is in S_{31} to D_1. The payload is an RLC of all the packets of A.
        if y==*32* **then**
            Send 1 packet whose index is in S_{32} to D_1. The payload is an RLC of all the packets of A.
        $x = x + 1$;
    end
else if $B == D_2$ **then**
    The same as above, replacing D_1 by D_2 and 11 by 22.

$$P_{grs11}(p_{11},\mathbf{l}^A,\mathbf{l}^B,\mathbf{u}_{ec})\ .$$

Algorithm 2: Protocol with IS-NC - Part 2

```
/* Else A ≠ S1 and A ≠ S2 and B ≠ D1 and B ≠ D2 */
else
    x = 0;
    while x ≤ w do
        Draw y from Hd(u).
        if y==11 then
            (1) Let n^11 be the list of indices in S11 at A that
            are neither in S11 nor in S31 at B (according to
            def. 3.1) and whose counter is strictly greater than
            1.
            Let p be the set of packets at A corresponding to
            these indices.
            if n^11 not empty and B − l^B ≥ 1 then
                Send a packet q to B with:
                  q.index = p1.index = n1^11
                  q.counter = ⌊p1.counter / 2⌋
                  q.payload = p1.payload
                Update
                  l11^B = l11^B + 1
                  p1.counter = ⌈p1.counter / 2⌉
                  Remove n1^11 from n^11

        if y==22 then
            Same steps from (1) as above, replacing 11 by 22
            and M by Q.

        if y==31 then
            (2) Let n^31 = [v, n^11], where n^11 stems from step
            (1) and v is the list of indices in S31 at A that are
            neither in S11 nor in S31 at B (according to def.
            3.1) and whose counter is strictly greater than 1.
            Let p be the set of packets at A corresponding to
            these indices.
            if n^31 not empty and B − l^B ≥ 1 then
                Send a packet q to B with:
                  q.index = p1.index = n1^31
                  q.counter = ⌊p1.counter / 2⌋
                  q.payload = RLC(all packets at A)
                Update
                  l31^B = l31^B + 1
                  p1.counter = ⌈p1.counter / 2⌉

        if y==32 then
            Same steps from (2) as above, replacing 11, 31
            and M by 22, 32 and Q.

        x = x + 1;
    end
    Go to the beginning of Algo. 2, exchange A and B and
    perform again all the steps.
```

$$dZ^1_{ic}(t) = N_c\sum_{e=1}^{C}\sum_{\mathbf{l}^A,\mathbf{l}^B}\beta_{ec}N_eP_{grs31}(i-l^B_{31},\mathbf{l}^A,\mathbf{l}^B,\mathbf{u}_{ec})\ldots$$

$$-N_c\sum_{e=1}^{C}\sum_{\substack{p_{31}>0,\mathbf{l}^A,\\ \mathbf{l}^B:l^B_{31}=i}}\beta_{ec}N_eP_{grs31}(p_{31},\mathbf{l}^A,\mathbf{l}^B,\mathbf{u}_{ec})\ .$$

The ODEs for Y_{ic} and Z^2_{ic} can be deduced from those of X_{ic} and Z^1_{ic}, replacing 1 by 2 everywhere. The components of the above equations are defined as follows:

- $P_j(c,\mathbf{l})$: fraction of relay nodes that are $(c,\mathbf{l})$-nodes. $P_j(c,\mathbf{l})$ is computed such that the following constraints are satisfied: $P_j(c,\mathbf{l})=0$ for $l^B>B$, $\sum_{\mathbf{l}}P_j(c,\mathbf{l})=1$, $\sum_{\mathbf{l}:l_{11}=i}P_j(c,\mathbf{l})=\frac{X_{ic}}{N_c}$, $\sum_{\mathbf{l}:l_{22}=i}P_j(c,\mathbf{l})=\frac{Y_{ic}}{N_c}$, $\sum_{\mathbf{l}:l_{31}=i}P_j(c,\mathbf{l})=\frac{Z^1_{ic}}{N_c}$ and $\sum_{\mathbf{l}:l_{32}=i}P_j(c,\mathbf{l})=\frac{Z^2_{ic}}{N_c}$.
- Let $K_{S_1}(t)$ be the number of indices released by S_1 up to time t and P_{sc} be the average number of indices that S_1 gives around time t to community c. Then, $\frac{dK_{S_1}(t)}{dt}=\sum_{c=1}^{C}\beta_{s_1c}N_cP_{sc}$ where, $P_{sc}=\sum_{p_{11}}\sum_{\mathbf{l}^B}p_{11}P_{gs11}(p_{11},\mathbf{l}^B,\mathbf{u}_{s_1c})P_j(c,\mathbf{l}^B)$. The number of indices of S_{11} that D_1 has received until time t is denoted by $R_{11}(t)$; $\frac{dR_{11}(t)}{dt}$ can be expressed from $P_{ls11}(.)$ in the same way as $K_{S_1}(t)$.
- $P_{nic,e,c}\big(n_{11},\mathbf{l}^A,\mathbf{l}^B,K_{S_1}(t),\mathbf{v}(t)\big)$: probability that for a $(e,\mathbf{l}^A)$-node and a $(c,\mathbf{l}^B)$-node, there are n_{11} indices of S_{11} at node A <u>n</u>ot <u>i</u>n <u>c</u>ommon with $S_{11}\bigcup S_{31}$ at node B and whose corresponding spray-counters are still below M, when S_1 has already spread out $K_{S_1}(t)$ indices. The vector $\mathbf{v}_{11}(c,t)$ stores the occurrence probability of each index of S_{11} at time t in community c: $\mathbf{v}_{11}(c,t)=\big(\frac{\tilde{X}_{1c}}{\tilde{X}_c},\ldots,\frac{\tilde{X}_{K_{S_1}(t)c}}{\tilde{X}_c}\big)$ with $\tilde{X}_c=\sum_{I\in S_{11}}\tilde{X}_{Ic}$. Similarly, we define $\mathbf{v}_{31}(c,t)$ from the $\tilde{Z}^1_{Ic}$ and we take $\mathbf{v}(t)=\frac{l^A_{11}}{s}\mathbf{v}_{11}(c,t)+\frac{l^A_{31}}{s}\mathbf{v}_{31}(c,t)+\frac{j}{s}\mathbf{v}_{11}(c,t)$, $s=l^B_{11}+l^B_{31}+j$ and $p_s=\frac{\sum_{I\epsilon E}\tilde{X}_{Ie}(t)}{\sum_{I\epsilon T_c}\tilde{X}_{Ie}(t)}$, with $T_e=\{I\in S_{11}:\tilde{X}_{Ie}>0\}$ and E be the set of indices of S_{11} that can still spread: $E=\{I\in S_{11}:0<\sum_{c=0}^{C}\big(\tilde{X}_{Ic}+\tilde{Z}^1_{Ic}\big)<M\}$. Then $P_{nic,e,c}\big(n_{11},\mathbf{l}^A,\mathbf{l}^B,K_{S_1}(t),\mathbf{v}(t)\big)$ and $P_{nicD,e,d_1}\big(n_{11},\mathbf{l}^A,\mathbf{l}^{D_1},K_{S_1}(t),\mathbf{v}(t)\big)$ (the probability that there are n_{11} indices of S_{11} at node B, not in common with S_{11} at D_1) are given by a combination of the above quantities with the function $S_z(.)$ defined hereafter. If S_e is a set of pairwise different elements from T_e whose cardinality is $|S_e|$, the probability to have exactly z different elements occurring among a set of $K_{S_1}(t)$ elements, the i^{th} elements having an occurrence probability $\mathbf{v}_i(t)$, is $S_z\big(K_{S_1}(t),z,\mathbf{v}(t)\big)=\sum_{S_e\subset T_e:|S_e|=z}\prod_{i\epsilon S_e}v_i(t)\prod_{i\epsilon T_e\setminus S_e}\big(1-v_i(t)\big)$. Further details of these derivations are given in the extended version of the paper [30].

In what follows, q_{11} $\big(Q_{11}$ for the random variable (r.v.)$\big)$ denotes the number of draws of S_{11} out of the bandwidth realization, n_{11} (N_{11} for the r.v.) denotes the number of indices of S_{11} that are in node A but not in B, p_{11} denotes the number of indices of S_{11} given by A to B, and s (ζ for the r.v.) denotes the bandwidth realization. Hence we have:

$$Pr(Q_{11}=q_{11})=\sum_{r\geq q_{11}}Pr(\zeta=r)\binom{s}{q_{11}}(u^{11}_{s_1c})^{q_{11}}(1-u^{11}_{s_1c})^{r-q_{11}}\ ,$$

$$Pr(N_{11}=n_{11}) = \begin{cases} P_{nicD,c,d_1}\big(n_{11},\mathbf{l}^B,\mathbf{l}^{D_1},K_{S_1}(t),\mathbf{v}(t)\big)\ , \\ \text{if } B=D_1 \\ P_{nic,e,c}\big(n_{11},\mathbf{l}^A,\mathbf{l}^B,K_{S_1}(t),\mathbf{v}(t)\big)\ , \\ \text{otherwise} \end{cases}$$

$Pr(\zeta = r)$ is given by the network configuration and can be any (taken as Poisson in the numerical examples below). Similar quantities are defined for S_{22}, S_{31} and S_{32}.

- $P_{gs11}(p_{11},\mathbf{l}^B,\mathbf{u}_{s_1c})$: probability that S_1 gives p_{11} indices (of S_{11}) to node B.
- $P_{ls11}(p_{11},\mathbf{l}^B,\mathbf{l}^{D_1},\mathbf{u}_{cd_1})$: probability that node B, upon meeting with D_1, drops p_{11} indices of S_{11} that D_1 already has or that B hands over to D_1. The derivations of both quantities above are quite straightforward and detailed in [30].
- $P_{grs31}(p_{31},\mathbf{l}^A,\mathbf{l}^B,\mathbf{u}_{ec})$: probability that node B receives p_{31} indices of S_{31} from node A.

$$P_{grs31}(p_{31},\mathbf{l}^A,\mathbf{l}^B,\mathbf{u}_{ec}) = (p_{31} \le B - l^B) \sum_{\substack{s,\\ n_{11},n_{31}}} F(p_{31})\dots$$

$$Pr(S=s)Pr(N_{11}=n_{11})Pr(N_{31}=n_{31})\dots$$

$$\Big(l^A_{31} > 0 \ or\ \big(l^A_{31}=0, l^A_{11}>0 \ and\ (l^A_{22}\ or\ l^A_{32}) > 0\big)\Big)\ ,$$

that represents the condition for generating a IS-NC packet (out of mixed S_{31} or unmixed packets in S_{11} and S_{22}). The r.v. S stands for the number of draws that elect a S_{11} or S_{31} index to be sent out, hence $Pr(S=s) = \sum_{r\ge s} Pr(\zeta=r)\binom{r}{s}(u_1)^s(1-u_1)^{r-s}$ with $u_1 = u^{11}_{s_{1c}} + u^{31}_{s_{1c}}$. Let $v_1 = u^{11}_{s_{1c}}/u_1$ and $v_3 = u^{31}_{s_{1c}}/u_1$. Specifically, $F(p_{31})$ captures the coding decision: when a contact occurs, at each transmission opportunity (below denoted by "a draw", the mean number of these being Bw), one of the four types of indices is drawn. If S_{31} is drawn, such an index is either directly one of the S_{31} indices at node A, or is one of the S_{11} if no S_{31} are yet available (the payload being forged by combining S_{11} and S_{22} or S_{32}). Hence, it leaves less S_{11} indices available for the subsequent draws of S_{11}. We have

$F(p_{31}) = Pr(p_{31}$ packets of S_{31}received in s draws$) = (p_{31} \le n_{31})\ f(p_{31},s) + (p_{31} > n_{31})$ $Pr\Big(n_{31}$ of S_{31} sent then $p_{31}-n_{31}$ sent from the S_{11} until $s\Big)$ with $f(p_{31}, s) = \binom{s}{p_{31}} v_3^{p_{31}} v_1^{s-p_{31}}$, and

$$Pr\Big(n_{31} \text{ of } S_{31} \text{ sent then } p_{31}-n_{31} \text{ sent from the } S_{11} \text{ until } s\Big) =$$

$$\sum_{a=n_{31}}^{s} Pr\Big(\text{all } n_{31} \text{ of } S_{31} \text{ exhausted at draw } a\Big)\dots$$

$$\sum_{b=a+p_{31}-n_{31}+1}^{\min(s,n_{31}+n_{11})} Pr\Big(\text{last packet in } S_{31} \text{ received at draw } b\Big)\dots \qquad (1)$$

$$Pr\Big(p_{31}-n_{31}\text{drawn in } b-a \text{ draws}\Big)\ ,$$

$$Pr\Big(\text{all } n_{31} \text{ of } S_{31} \text{ exhausted at draw } a\Big) = \binom{a}{n_{31}} v_3^{n_{31}} v_1^{a-n_{31}}\ ,$$

$Pr\Big($last packet in S_{31} received at draw $b\Big) = v_1^{s-b} + (b-n_{31} == n_{11}) - (b-n_{31} == n_{11})v_1^{s-b}$, and $Pr\Big(p_{31}-n_{31}$drawn in $b-a$ draws$\Big) = \binom{b-a}{p_{31}-n_{31}} v_3^{p_{31}-n_{31}} v_1^{b-a-(p_{31}-n_{31})}$.

- $P_{grs11}(p_{11},\mathbf{l}^A,\mathbf{l}^B,\mathbf{u}_{ec})$: probability that node B gains p_{11} indices of S_{11} from node A.

$$P_{grs11}(p_{11},\mathbf{l}^A,\mathbf{l}^B,\mathbf{u}_{ec}) = (p_{31} \le B - l^B)\dots$$

$$\begin{cases} \sum_{n_{11},q_{11}} Pr(Q_{11}=q_{11})Pr(N_{11}=n_{11})\dots \\ \big(p_{11} == \min(n_{11},q_{11})\big)\ \text{, if } l^A_{31}=0 \ and\ l^A_{22}=0 \\ \sum_{s,n_{11},n_{31}} Pr(S=s)Pr(N_{11}=n_{11})Pr(N_{31}=n_{31})\dots \\ G(p_{11}), \text{ otherwise} \end{cases}$$

with $G(p_{11}) = Pr(p_{11}$ packets of S_{11}received in s draws$) =$

$$\sum_{p_{31}}^{n_{31}+n_{11}-p_{11}} (p_{31} \le n_{31})A + (p_{31} > n_{31})B$$

with $A = Pr\Big(p_{31}$ of S_{31} drawn until s and p_{11} of S_{11} sent$\Big) = \binom{s}{p_{31}} v_3^{p_{31}}\ v_1^{s-p_{31}}\ \big((s-p_{31} > n_{31})(p_{31} == n_{31}) + (p_{31} \le n_{31})(s - p_{31} == n_{31})\big)$ and $B = Pr\Big(n_{31}$ of the n_{31} sent and then $p_{31}-n_{31}$ sent from the S_{11} and p_{11} of S_{11} sent until $s\Big)$ is given by eq. 1 with $Pr($last packet in S_{31} received at draw $b)$ changed to:

$Pr($last packet in S_{31} received at draw b and p_{11} sent in $s) = \big(a-n_{31}+b-a-(p_{31}-n_{31}) == n_{31}\big)\ Pr($all S_{11} exhausted in b draws$) + Pr($ no S_{31} drawn in $s-b\,)\ Pr($exactly $p_{11}-b+p_{31}$ sent between draws b and $s)$ the latter being obtained in a similar manner as for P_{grs31}.

4.2 Evolution of the index dissemination distribution

The ODEs for $\tilde{X}_{Ic}$ and $\tilde{Z}^1_{Ic}$ can be written as:

$$\frac{d\tilde{X}_{Ic}}{dt} = \sum_{e=1}^{C} \beta_{ce}N_eN_cA_{R11,e,c} + \beta_{s_1c}N_cA_{S11,c} - \beta_{cd_1}\tilde{X}_{Ic}A_{D11,c}\ .$$

$$\frac{d\tilde{Z}^1_{Ic}}{dt} = \sum_{e=1}^{C} \beta_{ec}N_eN_cA_{R31,e,c}\ .$$

The ODEs for $\tilde{Y}_{Ic}$ and $\tilde{Z}^2_{Ic}$ can be deduced from those of $\tilde{X}_{Ic}$ and $\tilde{Z}^1_{Ic}$, replacing 1 by 2 everywhere. We have the following components:

- $A_{D11,c}$: fraction of nodes in community c that have I of S_{11} in their buffer and that drop it upon meeting with D_1.
- $A_{S11,c}$: fraction of nodes in community c that are infected by S_{11}. The derivations of both quantities above are quite straightforward and detailed in [30].
- $pnth_{11,c}(I,\mathbf{l}^B)$: probability for node B not to have I of S_{11} in its buffer.

$$pnth_{11,c}(I,\mathbf{l}^B) = \frac{\sum_{j=l^B_{11}}^{B} S_{z,c}\big(K_{S_1}(t)-1, j, \mathbf{v}_{T_c-\{I\}}(t)\big)}{\sum_{j=l^B_{11}}^{B} S_{z,c}\big(K_{S_1}(t), j, \mathbf{v}(t)\big)}\ ,$$

where $\mathbf{v}(t) = \mathbf{v}_{11}(c,t)$. We define $pnth_{31,c}(I,\mathbf{l}^B)$ similarly, replacing $\mathbf{v}_{11}(c,t)$ by $\mathbf{v}_{31}(c,t)$.

- $A_{R11,e,c}$: fraction of nodes in community c without index I of S_{11} that obtain I from a relay in community e.

$A_{R11,e,c}$=

$$\begin{cases} 0 \quad \text{, if } \sum_{e=1}^{C}\big(\tilde{X}_{Ie} + \tilde{Z}^1_{Ie}\big) \ge M \\ \sum_{\mathbf{l}^A,\mathbf{l}^B} P_j(e,\mathbf{l}^A)P_j(c,\mathbf{l}^B)pnth_{11,c}(I,\mathbf{l}^B)pnth_{31,c}(I,\mathbf{l}^B)\dots \\ \big(1-pnth_{11,e}(I,\mathbf{l}^A)\big) \sum_{s,n_{11},n_{31}} Pr(S=s)Pr(N_{11}=n_{11})\dots \\ Pr(N_{31}=n_{31})\frac{p_{11}}{n_{11}}G(p_{11}), \text{ otherwise} \end{cases}$$

with $G(p_{11})$ is given in the above section. The term $\frac{p_{11}}{n_{11}}G(p_{11})$ is the probability that I is chosen to get forwarded given these conditions.

- $A_{R31,e,c}$: fraction of nodes in community c without index I of S_{31} that obtain I from another relay in community e. $A_{R31,e,c} =$

$$\begin{cases} 0 \quad , \text{if } \sum_{e=1}^{C} \left(\tilde{X}_{Ie} + \tilde{Z}_{Ie}^{1}\right) \geq M \\ \sum_{\mathbf{l}^A,\mathbf{l}^B} P_j(e,\mathbf{l}^A) P_j(c,\mathbf{l}^B) pnth_{11,c}(I,\mathbf{l}^B) pnth_{31,c}(I,\mathbf{l}^B) \dots \\ \Big((l_{31}^A > 0) A_R^{Case1} + \big((l_{31}^A = 0)(l_{11}^A > 0) \dots \\ (l_{22}^A \; or \; l_{32}^A) > 0\big)\big) A_R^{Case2} \Big) \text{ , otherwise .} \end{cases}$$

$A_R^{Case1} = (1 - pnth_{31,e}(I,\mathbf{l}^A)) pnth_{11,e}(I,\mathbf{l}^A) \sum_{\substack{p_{31} \leq B - l^B, \\ s, n_{31}, n_{11}}} (p31 \leq n31) \frac{p_{31}}{n_{31}} H + pnth_{31,e}(I,\mathbf{l}^A)(1 - pnth_{11,e}(I,\mathbf{l}^A)) \sum_{\substack{p_{31} \leq B - l^B, \\ s, n_{31}, n_{11}}} (p31 > n31) \frac{(p_{31} - n_{31})}{n_{11}} L$,

$$A_R^{Case2} = (1 - pnth_{11}(I,\mathbf{l}^A)) \sum_{\substack{p_{31} \leq B - l^B, \\ q_{31}, n_{11}}} \frac{p_{31}}{n_{11}} M \ .$$

The expressions of H, L and M are easily derived from the decomposition of P_{grs31} in the above section.

- $P_I^{11}(t)$: probability that D_1 has received index I of S_{11} by time t.

$$\frac{dP_I^{11}(t)}{dt} = \sum_{c=1}^{C} \beta_{cd_1} N_c \big(1 - P_I^{11}(t)\big) A'_{D11,c} \ ,$$

where $A'_{D11,c}$ is the fraction of nodes in community c that hold I of S_{11} and that hand I over to D_1 provided that D_1 does not have I. We refer to the extended version [30] for the detailed derivation of $A'_{D11,c}$.

Decoding Criterion: Let $P_{S_1}(\tau)$ be the success probability at time τ. To account for the possible benefit brought by coding while keeping a simple criterion, we consider that D_1 can recover the K_1 packets sent by S_1 if (i) it receives at least K_1 indices of S_{11}, or (ii) if it receives non-coded and coded packets so that all the K_1 and K_2 packets are received. Note that case (ii) is pessimistic as the coding matrix can be inverted even though it is not met, but it is so in order to keep a tractable decoding criterion. Yet, it allows to account for a coding benefit. Hence we have: $P_{S_1}(\tau)$=$P_{S_1}^{(i)}(\tau)$+$P_{S_1}^{(ii)}(\tau)$. The derivations of $P_{S_1}^{(i)}(\tau)$ and $P_{S_1}^{(ii)}(\tau)$ are given in [30].

4.3 Numerical validation

In this section, we assess the accuracy of the fluid model above, that captures the effect of the joint control of routing and IS-NC on various quantities. We consider a synthetic contact trace on which we run the IS-NC protocol described in Algo. 1-2 thanks to a discrete event simulator written in Matlab. The simulation results are averaged over 30 runs and the 5% confidence intervals are plotted. The trace is made of $N = 1000$ nodes, $C = 1$ for the sake of clarity of the curves and $\beta = 5.10^{-4}$. The buffer size is set to $B = 2$ packets. The bandwidth is Poisson distributed with mean $Bw = 3$ packets. The communication settings of the two sessions are: $K_1 = K_2 = 3$, $K_1^{'} = K_2^{'} = 5$ and $M = Q = 200$. We set the control policy $\mathbf{u}$ to $u_{11} = u_{22} = 0.3$, $u_{31} = 0.4$ and $u_{32} = 0$. Fig. 2 depicts the number of indices of each type packets, namely $\sum_{i=1}^{B} X_i$, $\sum_{i=1}^{B} Y_i$, $\sum_{i=1}^{B} Z_i^1$ and $\sum_{i=1}^{B} Z_i^2$. We observe the relative good fit between analysis and simulation for both non-coded and coded type packets. Fig. 3 represents the evolution of the normalized (i.e., divided by K_1 or K_2) number of DoFs of S_1 (resp. S_2) received by D_1. These numbers of DoFs are determined by a Gauss-Jordan elimination of the coding matrix in the simulation, and by the sum of the pairwise different received indices of S_{11} and S_{31} (resp. S_{22} and S_{32}) in the analytical model. We observe a good fit between the simulation results and the analytical prediction.

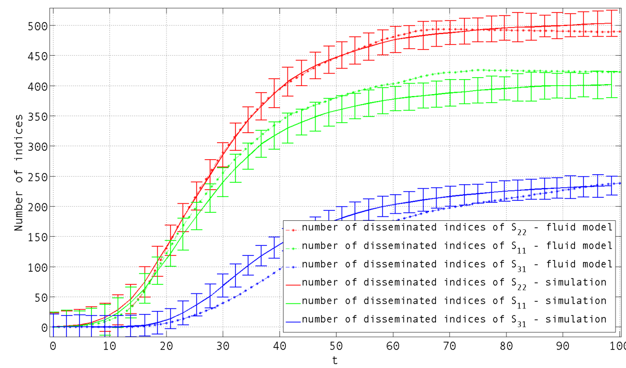

Figure 2: Evolution along time of the number of infected nodes with packets of different types.

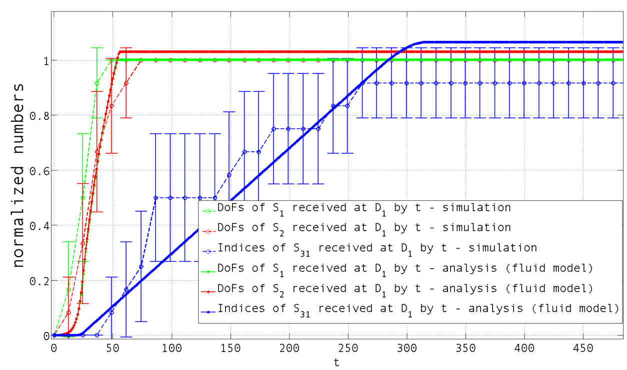

Figure 3: Evolution along time of the number of DoF of each source received by D_1.

5. THE IS-NC CONTROL PROBLEM

The problem we want to address thanks to (i) the introduced control policy of routing and IS-NC and (ii) the fluid model that predicts the delivery probability, is that of control policy optimization under some energy or memory constraint.

5.1 Discussion on the optimization problem

Let $U(.)$ be any classical utility function, such as $\log(1+x)$ if x is a probability, and $P_1(\tau)$ (resp. $P_2(\tau)$) the probability that D_1 (resp. D_2) has obtained its K_1 information packets by τ. The problem of finding the optimal policy $\mathbf{u}$ which jointly controls routing and pairwise IS-NC decisions under some energy constraint can be formulated as:

$$\max_{\{\mathbf{u_{ec}}\}_{\mathbf{e,c=1}}^{\mathbf{C}}} obj(\tau) = U(P_1(\tau)) + U(P_2(\tau))$$

subject to $\mathbf{u}$ satisfying the energy constraint.

Note that other objectives, such as the mean completion delay for each session, can be considered. Optimizing IS-NC decisions is a difficult problem in general, as discussed in Section 2, and in the social DTN scenario considered, this problem corresponds to a Markov decision process where at each time step, a central controller chooses an action so as to maximize the expected reward over a finite time horizon. It has been shown in [10] that when the system is made of N interacting objects and the occupancy measure is a Markov process (that has been discussed in Section 4), the optimal reward converges to the optimal reward of the mean field approximation of the system, which is given by the solution of an Hamilton-Jacobi-Bellman (HJB) equation.

Thanks to the ODEs presented in Section 4, that allow to get the fluid limits $P_{S_1}(\tau)$ and $P_{S_2}(\tau)$ of $P_1(\tau)$ and $P_2(\tau)$, the optimal IS-NC policy for a finite N can hence be approximated by the asymptotically optimal policy built by solving the HJB equation for the associated mean field limit. However, owing to the intricacy of our model made of coupled ODEs, the HJB equation cannot be solved in a closed-form. We would hence need to resort to a numerical solver, but the dimension of the involved vectors prevents from using this kind of solvers (see, e.g., [25]). A feasible implementation of the optimization procedure is to use heuristic optimization methods, such as Differential Evolution [27]. Besides, let us specify that in DTNs, a simple way of accounting for the energy consumption incurred by a routing policy is for example with the number of transmissions. This number can be easily extracted from the quantities modeled in Section 4, allowing to implement the energy constraint in the optimization process.

Investigating this optimization problem in social DTN thanks to the above fluid model is the subject of future work. In particular, in order to design a decentralized IS-NC policy, the model will be adapted to powerful existing decentralized routing policies (such as SimBet [6]) in order to devise relevant local IS-NC decisions.

5.2 Numerical example

We now provide a numerical example that shows the relevance of the approach trying to get benefit from IS-NC in social DTNs. We consider the topology depicted in Fig. 4 where the communities 1 and 3 are connected through another community 2. Community 1 (resp. 3) is that of the source node of session 1 (resp. 2) and of the destination node of session 2 (resp. 1) (these are 4 different nodes). This topology refers to the toy-example of two Wifi stations willing to exchange packets through an access point (AP) [16]. In this case, transmissions and hence time and throughput are saved if the AP combines the packets of the two stations. Whether this kind of IS-NC advantage can exist in DTNs is an open question, in particular with a policy taking decisions based on the community structure rather than on individual nodes, that is when the source and destination nodes of both sessions are not exchanged but represent four different nodes. On the simple topology of Fig. 4, we illustrate in Fig. 5, by simulations on a synthetic trace, that (community-based) IS-NC can be indeed beneficial with respect to intra-session NC. The intra-session NC policy we compare to is the best dissemination policy we found amongst those with varying values of $\mathbf{u}_{11}$ and $\mathbf{u}_{22}$ controlling spreading in community 1 and 3. Zhang et *al.* have shown empirically in [37] that uncontrolled IS-NC of different source-destination pairs is not beneficial in general in homogeneous DTNs. We illustrate in Fig. 5 that it can be beneficial in heterogeneous DTNs. Specifically, it turns out that that the highest benefit is obtained when session mixing is performed at the side communities, and not at the relay community, as the direct analogy with connected networks would suggest. Further study is needed, allowed by the presented protocol and its analytical model, to investigate on what social graph topologies (amongst which undirected like in Fig. 4) and under what conditions IS-NC can be beneficial.

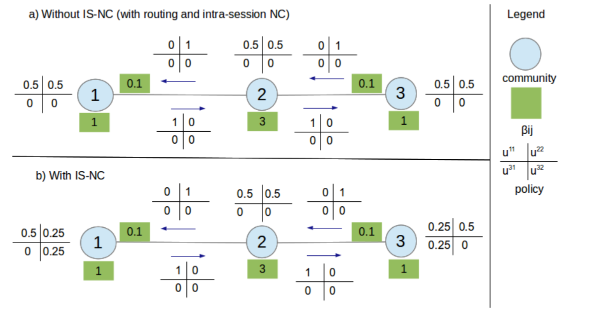

Figure 4: The community topology considered. Each community is made of 333 nodes and the inter-meeting intensities are in green. For both policies: $K_1 = K_2 = 7$, $K_1' = K_2' = 7$, $M = Q = 200$. (a) The non IS-NC policy (that is, a policy where only intra-session NC is used). (b) The IS-NC policy.

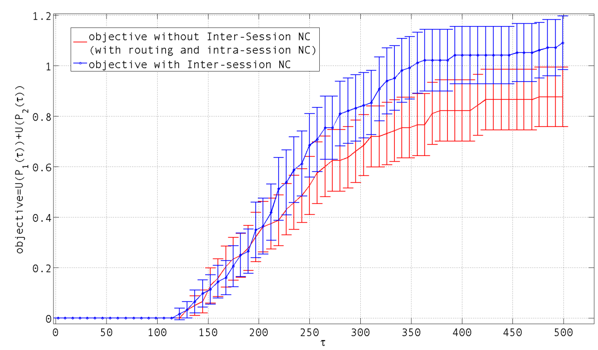

Figure 5: The obtained objective values for the above social DTN topology and policies.

5.3 Relevance to real-world traces

Finally, let us briefly show that even the simple topology of Fig. 4 can arise in real-world social DTNs. We consider as an example the MIT Reality Mining contact trace [7], corresponding to Bluetooth contacts collected with 100 smartphones distributed to students and staff at MIT over a period of 9 months. The people come into contact owing to the mobility and these contact patterns can reflect the social features such as the clustering of nodes into different communities, as analyzed in several studies such as [13]. We aggregated the contact trace into a weighted contact graph whose weights represent the tie strength (combining contact frequency and duration) between the nodes. We then applied Louvain community detection algorithm [3] to detect the communities in the contact graph and computed the β matrix describing the community structure. The following β matrix has been obtained with 7 communities detected: $\beta_{MIT} =$

$$\begin{bmatrix} 2.12 & 0.09 & 0.03 & 0.00 & 0.01 & 0.05 & 0.01 \\ 0.09 & 5.89 & 0.52 & 0.17 & 0.73 & 0.33 & 0.12 \\ 0.03 & 0.52 & 1.71 & 0.14 & 0.47 & 0.27 & 0.09 \\ 0.00 & 0.17 & 0.14 & 1.97 & 0.19 & 0.13 & 0.07 \\ 0.01 & 0.73 & 0.47 & 0.19 & 12.00 & 0.64 & 0.21 \\ 0.05 & 0.33 & 0.27 & 0.13 & 0.64 & 10.87 & 0.17 \\ 0.01 & 0.12 & 0.09 & 0.07 & 0.21 & 0.17 & 6.48 \end{bmatrix}.$$

A similar topology as in Fig. 4 arises when, for example, a node of community 1 and another node in community 4 want to exchange packets. In this case, a good route is to go through community 6, and the involved β_{ij} are then: $\beta_{11} = 2.1$, $\beta_{66} = 10.8$, $\beta_{44} = 1.9$, $\beta_{16} = 0.05$, $\beta_{64} = 0.13$.

6. CONCLUDING REMARKS

We have devised a parameterized pairwise IS-NC control policy and expressed the control optimization problem thanks to a performance model. The scheme is at the same time a routing and a coding policy that allows to optimize for a utility function defined over the two sessions, under some energy constraint. Our policy decides which nodes can mix the sessions based on their communities rather than on their individual properties, making the devised policy scalable with the number of nodes if the number of communities keeps limited. We have shown that numerical gains of IS-NC over intra-session NC can indeed be obtained.

This paper does not aim at presenting a self-contained decentralized IS-NC protocol that can be confronted with existing routing policies in DTN. It aims at devising, modeling and proving the benefit of a centralized social-aware (community-based) pairwise IS-NC policy. Specifically, the problem of grouping sessions by two is not investigated here. Detecting and selecting what pairs of sessions to be mixed is part of the decentralization problem. Moreover, further study is needed, allowed by the presented protocol and its analytical model, to investigate on what social graph topologies (amongst which undirected like in Fig. 4) and under what conditions (e.g., sizes K_1 and K_2 of the sessions and energy budget) IS-NC can be beneficial. The next step after this work is to study numerically the optimization problem in order to extract heuristics to devise a decentralized IS-NC policy for social DTNs. In particular, the model will be adapted to powerful existing decentralized routing policies (such as SimBet [6]) in order to devise relevant local IS-NC decisions.

7. ACKNOWLEDGEMENTS

This work has been funded by the French Research Agency under contract ANR-10-JCJC-0301.

8. REFERENCES

[1] R. Ahlswede, N. Cai, S. R. Li, and R. W. Yeung. Network information flow. *IEEE Trans. on Inf. Th.*, 46(4), Jul. 2000.

[2] S. Ahmed and S. S. Kanhere. Hubcode: hub-based forwarding using network coding in delay tolerant networks. *Wireless Comm. and Mob. Comp.*, 13(9):828–846, May 2011.

[3] V. D. Blonde, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *J. Stat. Mech.*, Jul. 2008.

[4] E. Bulut, W. Zijian, and B. Szymanski. Impact of social networks in delay tolerant routing. In *IEEE Globecom*, Honolulu, HI, Dec. 2009.

[5] P. Costa, C. Mascolo, M. Musolesi, and G. Picco. Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks. *IEEE JSAC*, 26(5), mar 2008.

[6] E. Daly and M. Haahr. Social network analysis for information flow in disconnected delay-tolerant MANET. *IEEE Trans. on Mob. Comp.*, 8(5):606 – 621, May 2009.

[7] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Pers. and Ubi. Comp.*, 10(4):255–268, Mar. 2006.

[8] A. Eryilmaz and D. Lun. Control for inter-session network coding. In *Inf. Theory and Appl. workshop (ITA)*, San Diego, CA, Jan. 2007.

[9] J. Fan, Y. Du, W. Gao, J. Chen, and Y. Sun. Geography-aware active data dissemination in mobile social networks. In *IEEE MASS*, pages 109–118, Nov. 2010.

[10] N. Gast, B. Gaujal, and J.-Y. Le Boudec. Mean field for markov decision processes: From discrete to continuous optimization. *IEEE Trans. on Automatic Control*, 57(9):2266–2280, Sep. 2012.

[11] R. Groenevelt and P. Nain. Message delay in MANETs. In *ACM SIGMETRICS*, pages 412–413, Banff, Canada, Jun. 2005.

[12] T. Ho, R. Koetter, M. Médard, D. R. Karger, and M. Effros. The benefits of coding over routing in a randomized setting. In *IEEE Int. Symp. on Info. Theory (ISIT)*, page 442, Yokohama, Japan, Jul. 2003.

[13] 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, 2011.

[14] 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*, pages 244–251, Aug. 2005.

[15] P. Hui, J. Crowcroft, and E. Yoneki. Bubble rap: Social-based forwarding in delay tolerant networks. *IEEE Trans. on Mob. Comp.*, 10(11):1576 – 1589, Nov. 2011.

[16] 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.*, 6(3):497 – 510, 2008.

[17] A. Khreishah, W. Chih-Chun, and N. Shroff. Rate control with pairwise intersession network coding. *IEEE/ACM Trans. on Netw.*, 18(3):816–829, Jun. 2010.

[18] R. Koetter and M. Medard. An algebraic approach to network coding. *IEEE/ACM Trans. on Netw.*, 11(5):782–795, 2003.

[19] T. G. Kurtz. Solutions of ordinary differential equations as limits of pure jump markov processes. *J. of Appl. Prob.*, 7(1), Apr. 1970.

[20] Z. Li and B. Li. Network coding in undirected networks. In *Annu. Conf. Inf. Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2004.

[21] Y. Lin, B. Li, and B. Liang. Efficient network coded data transmissions in disruption tolerant networks. In *IEEE INFOCOM*, pages 1508–1516, Phoenix, USA, Apr. 2008.

[22] Y. Lin, B. Li, and B. Liang. Stochastic analysis of network coding in epidemic routing. *IEEE JSAC*, 26(5), Jun. 2008.

[23] D. S. Lun, M. Medard, T. Ho, and R. Koetter. Network coding with a cost criterion. In *Int. Symp. on Info. Theory and its Appl. (ISITA)*, pages 1232–1237, 2004.

[24] S. Maheshwar, Z. Li, and B. Li. Bounding the coding advantage of combination network coding in undirected networks. *IEEE Trans. Inf. Theory*, 58(2):570– 584, 2012.

[25] I. M. Mitchell. A toolbox of level set methods, 2012.

[26] G. Neglia, X. Zhang, J. Kurose, D. Towsley, and H. Wang. On optimal packet routing in deterministic dtns. In *IEEE VTC*, Jun. 2013.

[27] K. Price and R. Storn. Differential Evolution (DE) for continuous function optimization.

[28] L. Sassatelli and M. Medard. Inter-session network coding in delay-tolerant networks under spray-and-wait routing. In *IEEE WiOpt*, pages 103–110, Paderborn, Germany, May 2012.

[29] N. Shrestha and L. Sassatelli. On optimality of routing policies in delay-tolerant mobile social networks. In *IEEE VTC*, Jun. 2013.

[30] N. Shrestha and L. Sassatelli. Extended version - http://www.i3s.unice.fr/tildesassatelli/MSextended2014.pdf, 2014.

[31] T. Spyropoulos, K. Psounis, and C. Raghavendra. Efficient routing in intermittently connected mobile networks: the multi-copy case. *IEEE/ACM Trans. Netw.*, 16(1):77–90, Feb. 2008.

[32] T. Spyropoulos, T. Turletti, and K. Obraczka. Routing in delay-tolerant networks comprising heterogeneous node populations. *IEEE Trans. on Mob. Comp.*, 8(8):1132 – 1147, Aug. 2009.

[33] D. Traskov, N. Ratnakar, D. Lun, R. Koetter, and M. Medard. Network coding for multiple unicasts: An approach based on linear optimization. In *IEEE Int. Symp. on Info. Theory (ISIT)*, pages 1758–1762, 2006.

[34] A. Vahdat and D. Becker. Epidemic routing for partially-connected ad hoc networks. In *Tech. Rep. CS-200006*, Duke University, 2000.

[35] J. Widmer and J.-Y. Le Boudec. Network coding for efficient communication in extreme networks. In *ACM SIGCOMM workshop on DTN*, pages 284–291, Philadelphia, USA, 2005.

[36] P. Yang and M. Chuah. Performance evaluations of data-centric information retrieval schemes for dtns. *ACM Comput. Netw.*, 53(4):541–555, mar 2009.

[37] 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.*, 21(5):1407–1420, Oct. 2013.