Joint Interest- and Locality-Aware Content Dissemination in Social Networks

Eva Jaho and Ioannis Stavrakakis
Department of Informatics and Telecommunications
National & Kapodistrian University of Athens
Ilissia, 157 84 Athens, Greece
Email: {ejaho, ioannis}@di.uoa.gr

Abstract—Social groups are typically formed by nodes that share common interests (interest-induced social groups), with no implication on the geographic location of these nodes. In addition to such groups, mobile nodes form groups also as they move around and come to a locality where they can establish communication with other nodes (locality-induced social groups). This paper investigates the intermingling of these distinct types of social groups and proposes an approach that can enhance content dissemination in the presence of such groups. Specifically, we introduce a framework for modelling this environment (both, the nodes' dynamic association to such groups and the dynamic *usability* of the content) and explore the conditions under which a proposed cooperative strategy can enhance the content dissemination process compared to a selfish one. This work basically explores how mobility and cooperative content storage strategies can help bridge interest-induced social groups, or how the joint association of nodes with interest- and locality-induced social groups can be exploited to enhance content dissemination.

I. INTRODUCTION

In social networks, nodes establish connections based on their social interests, forming interest-induced social groups. Such groups can generally be exploited in order to enhance the dissemination of information to participating nodes. For example, if a node is interested in music of the 80's, it may join the group "songs of the 80's" and allow all the members of this group to access the group's collective information content. This example can be applicable in publish/subscribe environments, in which the network delivers a published message only to the nodes whose subscribed interests match the content of the message [1], [2].

In environments where the information exchange is possible *only through opportunistic encounters* (which is the one considered in this paper), the benefits of interest-based social grouping can be harvested only by exploiting such encounters. The higher the chance for a node to encounter other nodes that are strongly associated with a certain

interest, the higher the probability for a node to acquire a certain content of interest.

Node encounters may occur in two ways:

- Locality-induced encounters: refer to encounters that occur in well defined localities which nodes have a good chance of visiting, based on their behaviour. Such localities will be considered to define locality-induced (as opposed to interest-induced) social groups. Such groups may be, for example, a coffee-break place, a train station, etc.

- Random encounters: refer to encounters not influenced by a particular locality (physical area) in which the node is found. For example, this may include encounters a node may have when moving between different locality-induced social groups.

In this paper, we explore how locality-induced node encounters and the nodes' own (content) interests can be jointly exploited to improve information dissemination in social networks.

In addition to its membership to interest-induced social groups, a node may be attributed membership to locality-induced social groups. Such memberships are expected to boost drastically the *discoverability* of (probability of acquiring) a desirable content. That is, the locality-induced social networking structure could direct the content dissemination process to target nodes which are likely to be encountered by nodes desiring the specific content. For example, nodes that are interested in music of the 80's may be members of the interest-induced social group "music of the 80's", but they could also be members of the locality-induced social group "discotheque XYZ", where they may frequently meet and thus, exchange more contents of the same interest. Such locality-induced social groups could be useful in enhancing content dissemination. In this paper we explore how mobility and cooperation can enhance content dissemination by exploiting the joint association of nodes with interest- and locality-induced social groups.

Interest- and locality-induced social grouping may be taken into account when a node decides on which type of content to store in its memory. Clearly, the type of content that each node stores in its memory and exchanges with other nodes, shapes the content dissemination pro-

cess. The effectiveness of this process can be measured in terms of a quantity (to be referred to as *valuability*) that captures both the *usability* and *discoverability* of the content; usability refers to the extent to which content is (still) useful (e.g., just created vs outdated and irrelevant any more, etc), while discoverability refers to the chance of succeeding in acquiring certain content by nodes that want it. This quantity will be clearly shaped by the adopted content dissemination strategy. The higher the value of this quantity, the more effective the adopted content dissemination strategy would be considered to be.

The paper is organized as follows. Section II refers to the related work. In Section III, we describe the attributes of each node that help define content dissemination strategies and the relevant attributes of a locality-induced social group. A mechanism for determining which content to store – which characterizes the content dissemination strategy – is described in Section IV. The effectiveness of the resulting content dissemination strategy is then determined by deriving the *valuability* of the contents under the considered content dissemination strategies in Section V. Section VI presents the performance evaluation results and finally, the paper is concluded in Section VII.

II. RELATED WORK

Exploiting interest-induced social groups has been shown to improve content dissemination in various networking environments. Consequently, detecting such groups is important and has received considerable attention. Example studies include the detection of communities in a web graph (where a community might correspond to sets of web sites dealing with related topics [3], [4]), detection of communities of scientists connected in a co-authorship graph [5], detection of communities in Delay Tolerant networking environments [6], etc. The dynamic detection of such groups of common interests, where users do not declare a priori their interests, has been studied in [7]. In that paper, the authors show that a proactive information dissemination within groups with common interests can reduce the search cost. Similarly, the authors in [8] show that detecting interest-based communities in Peer-to-Peer networks improves information dissemination and helps in pruning the search space.

The exploitation of both interest-induced social groups and locality-induced social groups in order to improve information dissemination, has recently attracted the attention of researchers. For instance, a dynamic scheme for deciding which objects (content) of a certain content type to replicate locally based on the encounters with other nodes is introduced in [9]. In that work each node appends a value to each object that is a function of its access probability and its availability in a locality, its size and the weight of the locality; this weight represents the relationship between the node and the locality (e.g., how often a node visits this locality). In [9] the objects' value does not change over time or space (where it is stored).

In our paper the problem is formulated differently than in [9]. The associations of each node with the interest- and locality-induced social groups are described through probability distributions over different content types (interests) and localities, which express the likelihood of a node to be interested in a certain content-type, or to visit a certain locality. Content exchanges are assumed to occur between a node visiting a locality-induced social group and the *entire* group, as a visit to a locality implies the ability to communicate with any member of that locality; as a result, exchanges are considered to be possible between the visiting node's storage and the collective storage of the group. The proposed content storage (and thus, dissemination) strategies operate on the content-type (or interest/content class) level and not on individual object level. The proposed cooperative strategy takes into consideration the interests of the locality-induced social groups the node is likely to visit in the future, thus aiming at serving as a bridge to distinct social groups and enhance content dissemination; this seemingly *altruistic* behaviour benefits the particular node as well, as the comparison results with a selfish counterpart to this strategy indicate. The proposed strategies are evaluated analytically with respect to a newly introduced performance metric called *valuability*; this quantity captures jointly how probable a certain content-type is to be found and how useful or usable it is (its *usability*). Among other factors, the *usability* may capture how fresh or novel an object is for a certain node (e.g., latest software update). Contents that resides outside a node's storage are considered to have high *usability* for this node, as such contents most likely have not been available to (or used by) that node in the past. After receiving such contents upon an encounter with another node, these contents are considered to become "old" as they are processed or utilized (if desirable) by the receiving node and, thus, their *usability* for the node that stores and carries them is considered to be decreased.

III. ATTRIBUTES OF SOCIAL GROUPS

In this section attributes of the interest- and locality-induced social groups are introduced that are used in the proposed framework for modelling these groups, as well as in the description of the investigated content storage and – ultimately – dissemination strategies.

Consider a social network that consists of N mobile nodes, C interest-induced social groups (or, equivalently, content classes) and L locality-induced social groups; the variables n, c and l will be used in the sequel to represent an element from these sets, respectively.

Network nodes are considered to belong in one or more interest-induced social groups. The degree of their association with each such group – that represents the distribution of content class preferences of the node – is captured by the node's self-interest factor. Let I_c^n, denote the self-interest factor of node n in content class c, where $0 \leqslant I_c^n \leqslant 1$ and $1 \leqslant c \leqslant C$. Let the self-interest vector of

node n be the collection of all self-interest factors of this node associated with all the content classes, denoted by $\overline{I}^n = (I_1^n, I_2^n, \ldots, I_C^n)$, where $\sum_{c=1}^{C} I_c^n = 1$.

In this work it is assumed that network nodes are mobile and exchange information only through encounters with other nodes. At any point of time these nodes may be found in a random location or in non-random, well-defined and somewhat popular locality; the latter are localities that are visited by a minimum number of nodes over certain periods and are, thus, considered to define a locality-induced social group. Let the self-movement factor of node n, M_l^n, be equal to the probability that node n is found in locality l at a random point in time, i.e., belongs in the locality-induced social group l, $0 \leqslant M_l^n \leqslant 1$. Let $\overline{M}^n = (M_1^n, M_2^n, \ldots, M_L^n)$, denote the self-movement vector of node n, containing the probabilities that node n is found in the various locality-induced social groups, where $\sum_{l=1}^{L} M_l^n \triangleq 1 - r^n$. r^n is the probability not to be in any of the locality-induced social groups. Without loss of generality and in order not to burden the presentation in this paper, we will consider that node encounters occur only within the defined localities (and not at random locations) and thus, $\sum_{l=1}^{L} M_l^n = 1$ or $r^n = 0$.

Let $\overline{A}^n = (\overline{I}^n; \overline{M}^n)$ denote the attributes of node n. As it will become clear later, the proposed cooperative content storage strategy will take into consideration certain attributes associated with the locality-induced social groups. Such attributes are defined in the sequel.

Let g_c^l denote the probability that a random member of group l belongs in content class c. This probability is given by

$$g_c^l = \sum_{n=1}^{N} P_l^n I_c^n, \tag{1}$$

where P_l^n denote the probability that a randomly selected node from locality-induced social group l is node n and is given by

$$P_l^n = \frac{M_l^n}{\sum_{j=1}^{N} M_l^j}.$$

Let w^l denote the weight of group l, defined to be equal to the average population of locality-induced social group l (mean number of nodes found in this group at a random inspection time), given by

$$w^l = \sum_{n=1}^{N} M_l^n. \tag{2}$$

Notice that w^l captures the ability of locality-induced social group l to diffuse content, as a higher value of w^l would suggest a higher potential for content dissemination within group l, since all the nodes within a locality-induced social group are considered to communicate with each others. For instance, if g_c^l is the same for all the locality-induced social groups l, it would be more effective to

try to forward contents of class c to the group with the largest population. An effective content storage (and, thus, content dissemination) strategy should aim at making content of class c available to locality-induced social groups that are well-populated by nodes utilizing this class of content as well as have a relatively large population.

IV. CONTENT STORAGE STRATEGIES

The above introduced framework for describing interest- and locality-induced social groups is employed in this section in order to define effective content storage strategies. These strategies will dictate the contents that the nodes will exchange upon encounters, so that (interest- and locality-induced) social grouping structure results in disseminating content effectively. Roughly speaking, an effective content dissemination strategy would increase the likelihood that a certain content type is made available to a node upon request. A specific metric for assessing the effectiveness of these strategies is presented in the next section.

In this paper – and in order to facilitate the presentation of the framework and show its potential for effectiveness – we restrict the consideration to the content (or interest) classes without getting into the fine resolution of the different objects within each content class. Consequently, the objective here is to devise a strategy that brings to the nodes the type of contents that are likely to be requested and not specific objects from a content class. In other words, the discussion is at the content-type level and not the object level. In addition to facilitating making the important points/contributions, this content-type level treatment could be almost directly applicable in an environment where there is an one to one equivalence between content-types and objects, while it is expected that the conclusions drawn here and the efficiencies achieved are also expected when considering metrics associated with the finer resolution of the object level.

As the nodes are assumed here to encounter other nodes within the specific localities and not in random locations, a node encounter in this paper is equivalent to an encounter of the node with an entire locality-induced social group. Consequently, the incoming node is expected to be presented with a (large) pool of contents and content-types that are stored in the storages of the nodes of this locality-induced social group. By assuming that the population of these groups is relatively large, it is reasonable to assume that all content types are in principle available, although not with the same number of objects. Consequently, the incoming node can find and place in its storage (that is considered to be extremely small compared to the total storage of all the nodes in the group) any type of content (although not necessarily any object) it desires. Which content-type to actually select is dictated by the employed content storage strategy.

Let IPI_c^n denote the *interest priority index* (IPI) of a node n for content class c, to be defined below. The content

storage strategies considered in this paper use this index to determine the portion of the node's storage that will be allocated for storing content of the different content classes when the node is confronted with such a decision (upon entering a locality-induced social group). That is, the proportion of storage each node n devotes to contents of class c is given by the normalized value of its IPI index,

$$\overline{IPI}_c^n = \frac{IPI_c^n}{\sum_{c=1}^{C} IPI_c^n}.$$

The first content storage strategy considered in this paper (to be referred to as the *selfish strategy*) utilizes the following IPI:

$$IPI_c^n = I_c^n. \tag{3}$$

That is, under this content storage strategy the nodes seek to store content-types that match completely their own content-type (self-interests), with no provision whatsoever for catering for the interests of other nodes they will encounter in the future.

The second content storage strategy considered in this paper (to be referred to as the *cooperative strategy*) utilizes the following IPI:

$$IPI_c^n = \sum_{l=1}^{L} M_l^n w^l g_c^l = \sum_{l=1}^{L} M_l^n \sum_{k=1}^{N} M_l^k I_c^k. \tag{4}$$

That is, under this content storage strategy the nodes seek to take into account the interests of the nodes they will most likely encounter in the future, aiming at maximizing the average benefit they can generate through such cooperative behaviour. The latter is achieved by defining the IPI by considering the locality-induced social groups the node is expected to visit (term M_l^n), the number of nodes within those groups (term w^l), and the interests of the nodes expected to visit those groups (term g_c^l).

V. Measuring the effectiveness of the content dissemination strategies

As indicated earlier and in order to show the potential effectiveness of the cooperative strategy without the exploding complexities of dealing with and keeping track of a realistically immense universe of individual objects, the discussions, definitions and overall evaluation will be confined to the content-type (as opposed to individual content) level. In view of the earlier discussions it is clear that the adopted content storage strategy will shape the effectiveness of the content dissemination process. The effectiveness of this process will be measured in this paper in terms of a quantity (to be referred to as *valuability*) that captures both the *discoverability* and the *usability* of the content.

Discoverability refers to the availability of a requested content-type c. This quantity will be clearly shaped by the adopted content dissemination strategy. The higher the value of this quantity, the more effective the adopted content dissemination strategy would *potentially* be, provided that the content that is made available to the node (discovered) is of high potential use, or *usability* as defined next.

Usability refers to the potential for making use of content that is, or becomes, available to a node (or ultimately, how useful a certain content is to the node). It is reasonable to assume that the content that a node finds in the storage of other nodes (upon entering a locality-induced social group) is of higher potential usage than the content that the node has been carrying in its own storage, as the latter content may be considered to have already been used by the node in the past, or be somewhat outdated.

Notice that it may be that the *discoverability* of a certain content may be high, but its *usability* be low and, consequently, a strategy that ignores the contents' *usability* would not be effective; such a strategy could end up providing contents to the nodes with high probability but these objects could be of small value to them. Clearly, if *discoverability* were the only criterion, the nodes would tend to bring to their own storage, contents of their own interests only (since the *discoverability* of contents stored locally is high), making them following in essence a "myopic" (or selfish) content storage behaviour. The latter would impact negatively on both the node itself (as it will likely end up carrying stale and largely useless content) and on a wider content dissemination (as it will not contribute to a wider circulation and refreshing of the content, benefiting all nodes). Through the introduced notion of *usability* we basically obtain a mechanism which, on one hand helps capture realistic aspects of the environment and avoid the above inefficiencies and, on the other hand, gives us a tool for stirring the behaviour of the nodes across the entire spectrum, from selfish (or even "pathologically" selfish) to cooperative (or even totally altruistic), depending on the *usability* values assigned to the contents. An effective content dissemination strategy would be the strategy that yields high content *valuability*, that is, taking into consideration both the *discoverability* and *usability* metrics.

Each node is assumed to assign a *usability* value to its contents of interest. It assigns a (relatively low) value v_l ($0 \leq v_l \leq 1$), to contents of its interest that are stored in its own storage, and it assigns a (relatively high) value v_r ($0 \leq v_r \leq 1$), to contents of its interest that are stored in other nodes. To keep the analysis simple, it is assumed that v_l and v_r maintain the same values for each node. The normalized *usability* values are given by $\overline{v_l} = \frac{v_l}{v_l + v_r}$ and $\overline{v_r} = \frac{v_r}{v_l + v_r}$, for v_l and v_r respectively.

The *discoverability* of a content-type can be expressed as the probability of acquiring content of this type. The probability that a node n finds contents of class c in its own (or, local) storage is given by

$$Pl_c^n = \overline{IPI}_c^n.$$

The probability that a node n finds contents of class c in the storage of other nodes (or, remote storage) is given by

$$Pr_c^n = \sum_{l=1}^{L} M_l^n [1 - \prod_{\substack{k=1, \\ k \neq n}}^{N} [M_l^k (1 - \overline{IPI}_c^k) + (1 - M_l^k)]].$$

That is, it is given by the sum for each locality-induced social group l of the probability that node n is in the group l, multiplied by the probability that at least one other node from this group has contents of class c.

The mean *valuability* of content class c for node n can now be defined by combining the above probabilities (capturing the *discoverability* of content) with the *usability* of contents, stored locally or remotely, as follows, for the case in which $\overline{v_r} \geq \overline{v_l}$:

$$V_c^n = \overline{v_r} Pr_c^n + \overline{v_l}(1 - Pr_c^n) Pl_c^n \tag{5}$$

When $\overline{v_r} \geq \overline{v_l}$, a node is primarily interested in acquiring contents that other nodes have (as they are considered to be more fresh and not exploited yet) with *usability* $\overline{v_r}$. Otherwise, if the contents of its interest are not found in other nodes, they are acquired from its local memory (in practice, they are kept and not moved out) with *usability* $\overline{v_l}$. That is, when $\overline{v_r} \geq \overline{v_l}$ node n first searches for contents of class c in other nodes (of higher *usability*) and if none is found, it searches its local memory. Notice that in this case the wider content dissemination is facilitated.

Note that if the notion of *usability* is absent, the mean *valuability* of content class c for node n can be calculated from the previous expression by setting $\overline{v_r} = \overline{v_l}$ and ignoring the resulting constant $\frac{1}{2}$, as it does not impact on the comparative study of the selfish and the cooperative content storage strategies investigated here.

By weighting (5) with the self-interest factors I_c^n of a node n for each content class $c = 1, ..., C$, the mean *valuability* of contents of its interest for node n, is derived by

$$V^n = \sum_{c=1}^{C} I_c^n V_c^n.$$

Finally, the mean *valuability* of contents of its interest for a random node, is given by

$$V = \frac{\sum_{n=1}^{N} V^n}{N}. \tag{6}$$

In the next section we will investigate the effectiveness of the selfish and cooperative content storage (or dissemination) strategies.

VI. NUMERICAL EVALUATION

In this section we derive results on the mean *valuability* of contents of interest for a random node (as expressed in (6)) under the selfish and cooperative content dissemination strategies introduced earlier. The higher the value of *valuability*, the more effective the associated content dissemination strategy is considered to be, according the earlier discussions. The focus of the study is on cases where nodes have different preferences for content classes, so that the mutual benefit of cooperation be revealed. In all cases considered here, all nodes are either selfish or all are cooperative.

The self-interest vector $\overline{I}^n = (I_1^n, I_2^n, \ldots, I_C^n)$ is drawn from a Zipf distribution with exponent $s_C = 5$. The probability that each node is interested in contents of class c, $c = 1, ..., C$, which is the self-interest factor of this node in content class c, is given by

$$I_c^n = f(c; s_C, C) = \frac{1/c^{s_C}}{\sum_{k=1}^{C} 1/k^{s_C}}. \tag{7}$$

The drawn probabilities are the same for all the nodes and they result in a preference ranking for the C content classes, which are accordingly ordered as $[1, 2, ..., C]$. Since we are interested in considering the case where each node has different ordering of preferences for each content class, node 1 is assigned the preference order $[1, 2, ..., C]$ and this order set is shifted to the left by $n-1$ positions to generate the preference ranking for node n ($n = 2, ..., N$). Thus, if the preference ranking of node 1 is $[1, 2, ..., C]$, then the ranking of node 2 is $[2, 3, ..., C, 1]$, the ranking of node 3 is $[3, 4, ..., C, 1, 2]$, and so on.

The self-movement vector $\overline{M}^n = (M_1^n, M_2^n, \ldots, M_L^n)$ is drawn from a Zipf distribution (that is, from (7) by substituting I_c^n with M_l^n, s_C with s_L and C with L) with exponent s_L varying from $s_L = 0$ (in which case the distribution of localities is uniform and the mobility of the nodes is considered to be high as the nodes visit with the same probability all the locality-induced social groups) to $s_L = 5$ (the mobility of the nodes is considered to be low). Again, we consider the case where the preference ranking of node 1 is $[1, 2, ..., L]$ for the L localities, the ranking of node 2 is $[2, 3, ..., L, 1]$, the ranking of node 3 is $[3, 4, ..., L, 1, 2]$, and so on.

The values of g_c^l and w^l are calculated from (1) and (2), respectively.

A. Results for various self-movement vectors of a node

We first consider the case for $N = C = L = 3$. In this case and in view of the earlier assumptions on the preference ranking, it turns out that all three nodes have different largest preferences for content classes and localities (for $s_C = 5$ and $s_L > 0$). To show the impact of *usability* on the mean *valuability* of contents of interest for a random node (and, thus, on the effectiveness of the content dissemination process), we consider the cases where $\frac{v_r}{v_l} = 1, 5, 10$, or 100.

Fig. 1 presents results for the different values of $\frac{v_r}{v_l}$. These results show that the selfish strategy outperforms the cooperative one for any value of the parameter s_L
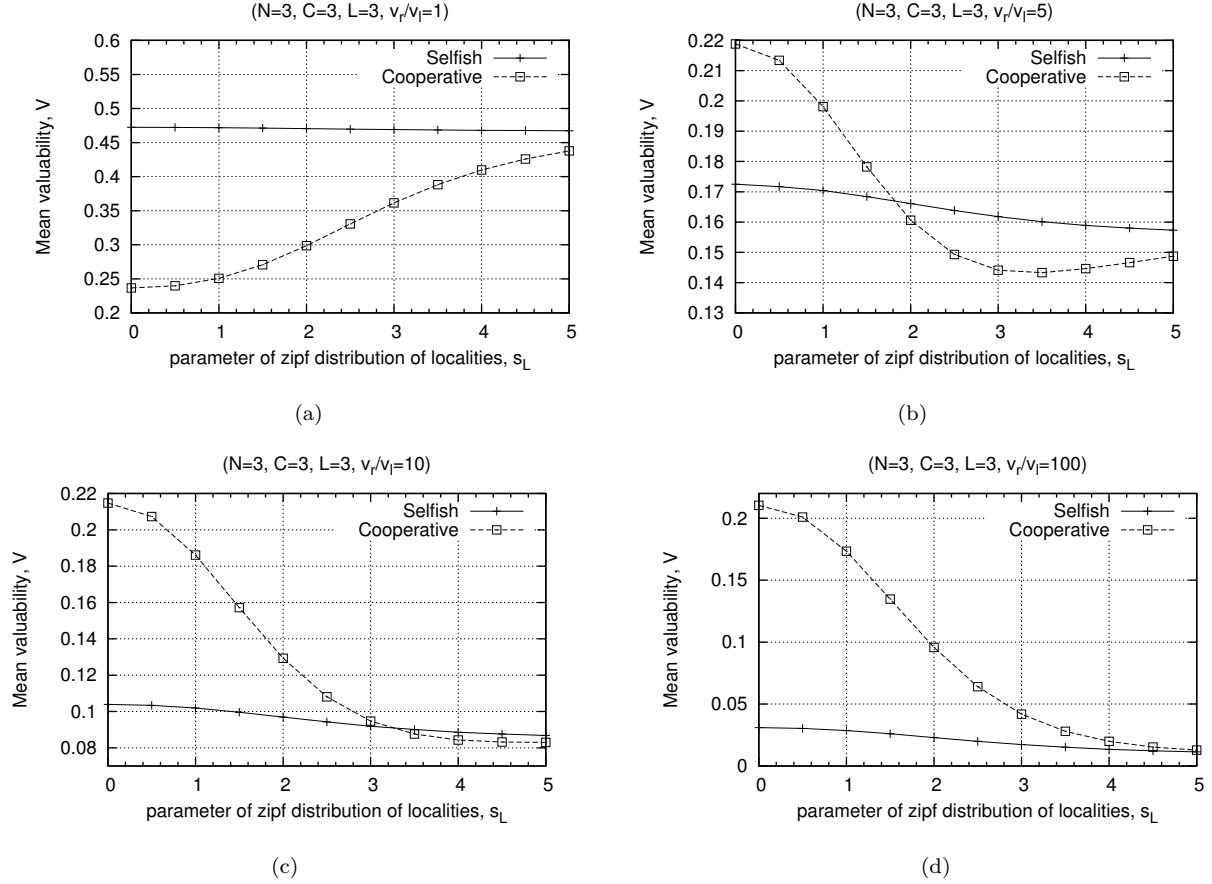
Fig. 1. Mean *valuability* of contents of interest for a random node for different values of parameter s_L

when all the contents are considered to have the same *usability*, or equivalently, when the nodes ignore it (as shown in Fig. 1(a)). When mobility is high ($s_L = 0$), the cooperative strategy is very inefficient, since the nodes have different preferences for content classes and thus, the probability of finding contents of interest in other nodes is low. When the nodes have high probability of visiting certain localities only and much smaller probability of visiting others ($s_L = 5$) the performance of the cooperative strategy approaches that of the selfish. This is due to the fact that the IPI value under the cooperative strategy (as calculated from (4)) approaches that under the selfish strategy (as calculated from (3)), and thus, the probability of finding contents of interest in other nodes and locally becomes almost the same.

Fig. 1(b), 1(c) and 1(d) show results for various values of $\frac{v_r}{v_l} > 1$. Under high node mobility ($s_L = 0$), which can be expressed with a uniform distribution of localities ($s_L = 0$), the cooperative strategy outperforms the selfish one, clearly indicating that mobility bridges nodes with different preferences for content classes, as long as the *usability* of the remote contents is larger than that of the locally stored ones. As s_L increases (and the nodes have a higher probability of staying at a certain locality), the

difference between the cooperative and the selfish strategy decreases. The cooperative strategy performs generally well, since it outperforms the selfish one for high mobility and it approaches the selfish one for very low mobility. Under medium mobility, depending on the magnitude of the ratio $\frac{v_r}{v_l} > 1$, the cooperative strategy may outperform the selfish one (for very high values of this ratio, e.g., $\frac{v_r}{v_l} = 100$) or not (e.g. for $\frac{v_r}{v_l} = 10$).

The mean *valuability* under the selfish strategy decreases as s_L increases, since the nodes tend to stay in different localities and thus, the probability of finding contents of interest in other nodes decreases.

As the value of $\frac{v_r}{v_l}$ increases, suggesting that contents of other nodes are more important, the cooperative strategy outperforms the selfish one under a wider ranges of values of s_L. More specifically, in Fig. 1(d), where $\frac{v_r}{v_l} = 100$, the cooperative strategy outperforms the selfish one for all the values of s_L as compared to the Fig. 1(b) and 1(c) where the value of $\frac{v_r}{v_l}$ is lower and thus, the cooperative strategy outperforms the selfish one only for some values of s_L. Clearly, there is a threshold of the ratio v_r/v_l above which the higher *usability* value of the remote contents more than compensates for the lower *discoverability* of the remote contents, making the *valuability* of the cooperative
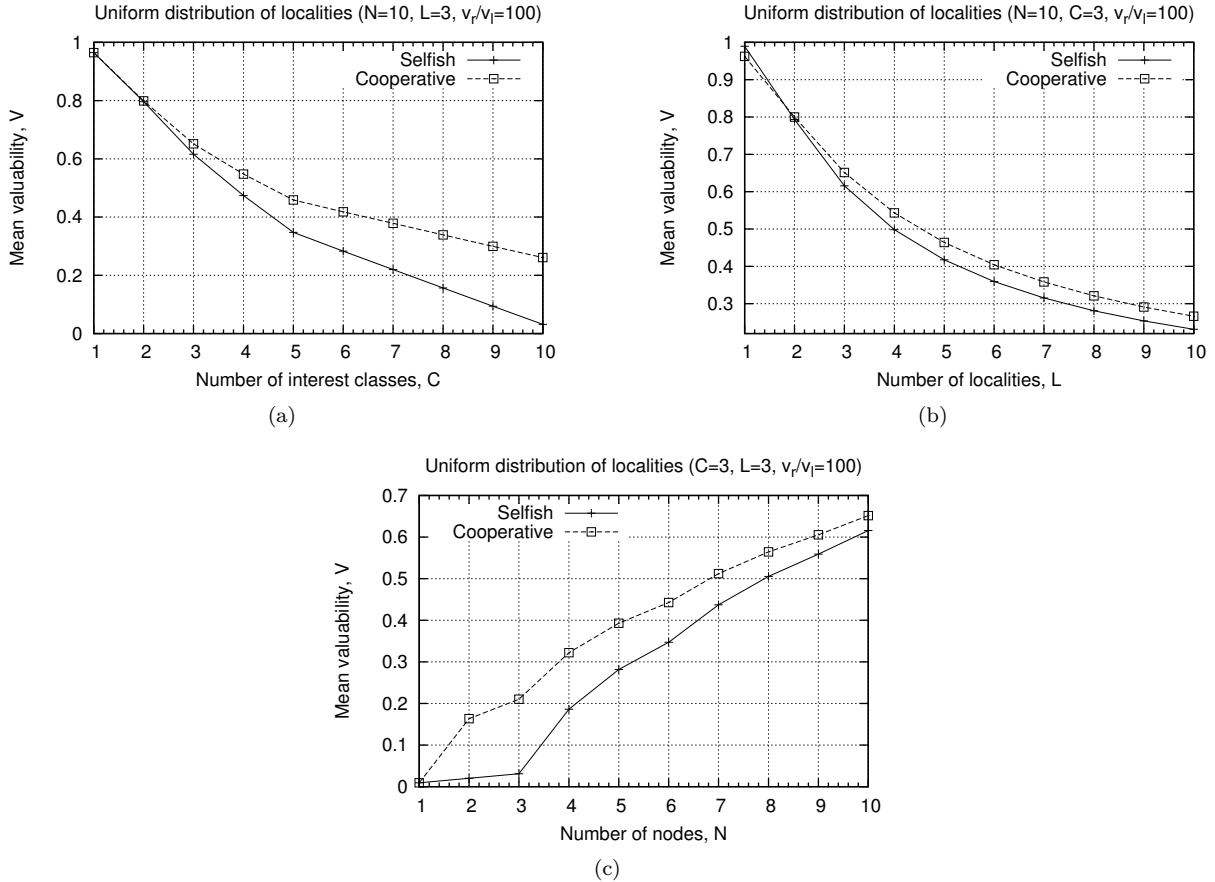
Fig. 2. Mean *valuability* of contents of interest for a random node for uniform distribution of localities

strategy (which stores remote contents) higher than that of the selfish one.

B. Results for identical self-movement factors

Fig. 1 clearly demonstrates that under high mobility ($s_L \simeq 0$), both the selfish and the cooperative strategies perform well. In addition, for a high ratio of $\frac{v_r}{v_l}$, for example $\frac{v_r}{v_l} = 100$, when $0 \leq s_L \leq 5$, the cooperative strategy always outperforms the selfish one. For this reason, we focus in this subsection on the case of high mobility; notice that in this case, the nodes have equal self-movement factors, or equivalently, a uniform distribution of locality preferences. Specifically, the case of $s_L = 0$ and $\frac{v_r}{v_l} = 100$ is considered in this subsection and the performance of the two strategies is investigated as the values of C, L and N vary; the parameter s_C is fixed at $s_C = 5$. Results are shown in Fig. 2.

Fig. 2(a) and 2(b) show that when each node visits each locality with the same probability, the selfish and cooperative strategies perform similarly for small values of C and L. This suggests that mobility can help in the dissemination of contents, even when nodes are selfish, provided that the number of content and localities is small.

Fig. 2(a) results for the case of $N = 10$, $L = 3$ and C varying from 1 to 10. When there is only one content class ($C = 1$), V is the same for both strategies and almost equal to 1. This result is anticipated, since each node has contents of only one class and thus, a cooperative behaviour would also serve a node's own interests. As C increases the difference between selfish and cooperative strategies becomes more pronounced, with the latter outperforming the former. This shows that cooperation is more beneficial than selfishness if there exist more than one content classes. Moreover, V decreases for both the selfish and cooperative strategies as C increases. This is attributed to the fact that the nodes have different preference rankings for content classes and, thus, the more the content classes, the lower the probability for a node of finding contents of its interest in other nodes, and thus, the lower the value of V.

Figure 2(b) shows the same mean *valuability* V as a function of the total number of locality-induced social groups L, which varies from 1 to 10. There are $N = 10$ nodes and $C = 3$ content classes. Since each node n ($1 \leq n \leq 10$) has a different preference ranking for the three content classes, by shifting the preference order set $[1, 2, 3]$ to the left by $n - 1$ positions, there would be four nodes with the same preference ranking $[1, 2, 3]$, three nodes with the ranking $[2,3,1]$, and three other nodes with

the ranking $[3, 1, 2]$. If there is only one locality $(L = 1)$, a selfish behaviour would be better than a cooperative one, since in the same locality it is highly probable that there can be other nodes with the same preference ranking for the content classes. However, as L increases, cooperation clearly outperforms selfishness.

Figure 2(c) shows V as a function of the total number of nodes, N. It is anticipated that as N increases, the mean *valuability* V increases as well, since the probability of finding contents of interest in other nodes increases. Also, as $N \leq C$ the selfish strategy for all nodes is much worse than the cooperative one. That is due to the fact that because the nodes' preference rankings for content classes do not coincide and thus, selfishness is in such a case detrimental to content dissemination. As N increases (for $N > C$) – keeping the same number of localities L and content classes C – the difference in performance between the cooperative and selfish strategies is much smaller, since there are nodes in the same localities with the same preference ranking for the content classes.

VII. Conclusion

In this paper we introduced a framework for modelling and disseminating content in a networking environment comprised of interest- and locality-induced social groups. Node encounters – or node visits to locality-induced social groups – are considered to be the mechanism for content exchanges and ultimately content dissemination. Additional specific contributions of this work are the following.

Two content storage or content dissemination strategies have been introduced and investigated: the selfish and the cooperative ones. Selfish nodes store locally only contents of their interests, whereas cooperative nodes also consider the interests of other nodes that are likely to be encountered in the future when deciding on the content to store locally.

A new metric for measuring the effectiveness of content dissemination is introduced, the content *valuability*, that takes into account both *usability* and *discoverability* of the content. The *usability* represents the degree of potential usefulness or usage of certain content, capturing the fact that not all content is of the same value. Specifically it has been argued that a high value of *usability* should be attributed to contents of interest stored in the other nodes whereas, a low value should be attributed to locally stored contents, since they are considered to have already been exploited or be outdated and be of lesser value. The *discoverability* of the content represents the probability of acquiring certain content.

By employing the aforementioned metric we investigated the performance of the two strategies and explored the conditions under which the cooperative strategy can enhance the content dissemination process compared to a selfish one. A number of results have been derived showing the comparative performance of the two strategies. Results have shown that the cooperative strategy outperforms the selfish one under high mobility, whereas, it has a similar performance in cases where nodes have high probability of staying in certain localities. Under identical probabilities of visiting any locality-induced social group, when the total number of localities and content classes increase, the difference in performance between the cooperative and selfish strategy becomes larger with the former having a better performance.

Concluding, in this paper we have explored how mobility and cooperation can enhance content dissemination by exploiting the joint association of nodes with interest- and locality-induced social groups.

References

[1] E. Yoneki, P. Hui, S. Chan, and J. Crowcroft, "A socio-aware overlay for publish/subscribe communication in delay tolerant networks." in *MSWiM*. ACM, 2007, pp. 225–234.

[2] G. Chockler, R. Melamed, Y. Tock, and R. Vitenberg, "Spidercast: a scalable interest-aware overlay for topic-based pub/sub communication," in *DEBS '07: Proceedings of the 2007 inaugural international conference on Distributed event-based systems.* New York, NY, USA: ACM, pp. 14–25.

[3] G. W. Flake, S. Lawrence, C. L. Giles, and F. M. Coetzee, "Self-organization and identification of web communities," *Computer*, vol. 35, no. 3, pp. 66–71, 2002.

[4] G. Flake, S. Lawrence, and L. L. Giles, "Efficient identification of web communities," in *Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Boston, MA, March 2000, pp. 150–160.

[5] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proc Natl Acad Sci U S A*, vol. 99, no. 12, pp. 7821–7826, June 2002.

[6] P. Hui, E. Yoneki, S.-Y. Chan, and J. Crowcroft, "Distributed community detection in delay tolerant networks," in *Proc. MobiArch*, 2007.

[7] A. Iamnitchi and I. Foster, "Interest-aware information dissemination in small-world communities," in *HPDC '05: Proceedings of the High Performance Distributed Computing, 2005. HPDC-14. Proceedings. 14th IEEE International Symposium.* Washington, DC, USA: IEEE Computer Society, 2005, pp. 167–175.

[8] K. D. D. P. Khambatti, M. Ryu, "Structuring peer-to-peer networks using interest-based communities," *Lecture Notes in Computer Science*, vol. 2944, pp. 48–63, September 2004.

[9] C. Boldrini, M. Conti, and A. Passarella, "Contentplace: Social-aware data dissemination in opportunistic networks," in *The 11-th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM'08)*, 2008.