Social Similarity as a Driver for Selfish, Cooperative and Altruistic Behavior

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

***Abstract*—This paper explores how the degree of similarity within a social group can be exploited in order to dictate the behavior of the individual nodes, so as to best accommodate the typically non-coinciding individual and social benefit maximization. More specifically, this paper investigates the impact of social similarity on the effectiveness of content dissemination, as implemented through three classes representing well the spectrum of behavior-shaped content storage strategies: the selfish, the self-aware cooperative and the optimally altruistic ones. This study shows that when the social group is tight (high degree of similarity), the optimally altruistic behavior yields the best performance for *both* the entire group (by definition) and the individual nodes (contrary to typical expectations). When the group is made up of foreigners with almost no similarity, altruism or cooperation cannot bring much benefits to *either the group or the individuals* and thus, a selfish behavior would make sense due to its simplicity. Finally, the self-aware cooperative behavior could be adopted as an easy to implement distributed scheme – compared to the optimally altruistic one – that has close to the optimal performance for tight social groups, and has the additional advantage of not allowing mistreatment to any node (i.e., the content retrieval cost become larger compared to the cost of the selfish strategy).**

I. Introduction

Today's networks can be highly personalized, in the sense that their structure and usage are shaped by the personal interests and behavior in general of the participating nodes. Nodes in such networks – referred to as social networks – are typically well connected, develop reciprocal trust relations, and have some common features, such as the content they are interested in and the places they tend to visit. Groups of such nodes are called social groups [11].

In this paper, we consider a group of (networked) nodes with common interests in content – more generally: objects; in computer science objects of interest are usually information objects, such as files and software. It is assumed that nodes of this social group store objects in their limited local storage to retrieve them when desired at minimum cost. If the nodes do not possess a desired object, they can fetch it either from a node in the group at some low-medium cost or – if not available in the group – from a node outside the group at higher cost. The low-medium cost associated with fetching an object from within the group may reflect actual or virtual price, access delay due to the locality of the fetching process or level of connectivity, and level of trust and cooperation.

As the local storage is assumed to be limited when compared with the plethora of objects possibly desired, an inherently selfish node would tend to store locally objects of higher personal interest. Our past work in [8] has shown that this is not the best content placement strategy for a node in a distributed group with the three levels of content access cost considered here. Instead, a cooperative content placement strategy has been devised based on game-theoretic arguments. The cooperative strategy determines which objects each node should store locally, so that the total content access cost for each and every node is no more than (and typically much lower than) that induced under the selfish strategy. The latter property implies that the content placement strategy is *mistreatment-free*: no node will lose by participating in the group, compared to acting selfishly. Hereafter we will refer to this placement strategy as the self-aware cooperative strategy, due to its mistreatment-free property and cooperative nature.

Mistreatment-free strategies are key to the sustainability of such distributed selfish groups, as they motivate users to participate in the group and share objects with others. The social benefit (i.e., the average benefit over all nodes) induced by the self-aware cooperative strategy is not optimal. The strategy that implements a content assignment that maximizes the social benefit will be referred to hereafter as the optimally altruistic strategy; this (optimal) content assignment can be derived by solving an optimization problem (as done, for instance, in [9]). The implementation of the optimally altruistic strategy would require the exchange of richer information among the nodes in the group (local demand distributions), whereas the self-aware cooperative strategy requires the exchange of some limited information among the nodes in the group (indices of content stored locally). Finally, note that to maximize the total benefit, some of the nodes may end up gaining too much and others being mistreated.

Focus of this paper: It is, therefore, evident that a node participating in a distributed group may face a dilemma as to which strategy, to follow. In this paper the characteristics of the social group are exploited in order to help address the above dilemma. We propose an innovative approach to

This work has been supported by the IST-FET project SOCIALNETS (FP7-IST-217141).

978-1-4244-7265-9/10/$26.00 2010 IEEE

characterizing the similarity of the social group nodes with respect to content interests and introduce a group *tightness* metric. The dependence of the induced social and individual node benefits on the level of *tightness* of the social group is clearly established for all three aforementioned content placement strategies, which reflect general patterns of social behavior. Our results let us draw important conclusions and guidelines for the content placement strategy a node should adopt for a given level of *tightness* in the social group.

Related work:The exploitation of social characteristics for data dissemination in autonomic and opportunistic networks has been considered from various aspects, in the literature. In [1], the authors construct a dynamic learning algorithm where nodes from various social communities opt for a utility-maximizing content placement strategy based on their encounters with other nodes. [14] studies the impact of different "levels of altruism" of nodes involved in the dissemination process. In a more abstract setting, the effect that a node's relational position in the group has on content dissemination has been considered in [4]. The importance of designing socially-aware opportunistic networks is also demonstrated in [2], [5], [7]. Finally, for a review of data dissemination in the general context of opportunistic networks, readers are referred to [3].

This study applies to social networks with interactions between computer devices having limited memory resources. These are typically encountered in mobile opportunistic networks that are additionally "socially aware", meaning that either the nodes or their human users are aware of the formation of social groups and the potential benefits from participation in such a group.

In Section II we formulate the problem and introduce the tightness metric capturing the degree of similarity of interests within a social group. In Section III the three strategies for content placement are briefly described. These strategies are compared to eachother in section IV, under different *tightness* values, with respect to the induced content retrieval cost at both the individual node and entire group level. Finally, we summarize the major conclusions of the paper in Section V and point to interesting problems for future work.

II. Problem Formulation And Tightness Metric

We assume that there are N nodes in a social group and each node has its own probability distribution of interest in M information objects (preferences). Let $\mathcal{M} = \{1, 2, ..., M\}$ be the set of objects, $\mathcal{N} = \{1, 2, ..., N\}$ be the set of nodes and F^n be the interest distribution of node n over the objects. F^n_m can also be viewed as the request rate of node n, $(n = 1, ..., N)$ for object m, $(m = 1, ..., M)$. All objects are assumed to be unit-sized. Node n has a storage capacity of C_n units.

Let P_n denote the *placement* of node n, defined to be the set of objects stored locally at this node. Without loss of generality, we take $|P_n| = C_n$ since a node can always gain by saving objects of interest locally in its storage than having to retrieve them from a distant source. Let $P = \{P_1, P_2, \ldots, P_N\}$ denote the global placement for the social group and $P_{-n} = P \setminus P_n$ the set that contains the placements of all nodes in the group except for node n.

Assume that the cost for accessing an object from a node's local storage is t_l, from another remote node in the group t_r and from another node in another group t_s, with $t_l < t_r < t_s$. These values are assumed to be the same for all nodes in order to simplify the analysis. In reality, the three cost types may be different for each node; yet the costs for different node pairs within a social group are expected to be similar.

Given an object placement P, the mean access cost per unit time for node n is given by:

$$\mathcal{C}_n(P) = \sum_{m \in P_n} F^n_m t_l + \sum_{\substack{m \notin P_n, \\ m \in P_{-n}}} F^n_m t_r + \sum_{\substack{m \notin P_n, \\ m \notin P_{-n}}} F^n_m t_s \, . \quad (1)$$

The first summation corresponds to the mean cost of accessing objects locally; the second term refers to the mean cost of accessing them from nodes within the social group; and the third sum accounts for the mean cost of accessing objects not stored anywhere in the group, i.e., from a node external to the group.

To define *tightness* as a measure of similarity between the nodes' preferences for objects, we used the Kullback- Leibler (K-L) divergence [6], a well-known metric capturing the divergence between two distributions. The Kullback-Leibler divergence of distribution Q from S is defined as:

$$D_{S,Q} = \sum_i S(i) log \frac{S(i)}{Q(i)}.$$

Besides being always non-negative, K-L has some desirable properties in the considered context of the paper that favor it over other well-known (dis)similarity metrics, such as the Kolmogorov-Smirnov distance Spearman's rank correlation coefficient, proportional similarity, and total variation distance [10], [13]. As opposed to the Kolmogorov-Smirnov distance which takes the supremum of the differences over all elements of a distribution, in the K-L divergence all differences contribute to the calculation. Thus, the number of preferences that two nodes have into account is also implicitly considered. Spearman's rank correlation coefficient describes the degree of association in the ranking of the distribution elements; hence, it would not provide us with insights when the ranking of interest in objects is the same between two nodes but the actual distribution values are different. Finally, the K-L divergence permits our metric of tightness to take a broader range of values compared to the proportional similarity or total variation distance. Our comparative evaluations suggest that K-L is more sensitive to changes of distribution values, and thus can more accurately depict differences in interest profiles.

K-L is not a measure of distance since $D_{S,Q} \neq D_{Q,S}$. To come up with one, we invoke the symmetrized divergence, which is defined as:

$$\overline{D}_{S,Q} = \overline{D}_{Q,S} = D_{S,Q} + D_{Q,S}.$$

Hence, we can define $\overline{D}_{F^i,F^j}$ as the distance of preference distributions of nodes i and j (to be referred to hereafter as the preference distance between i and j). Notice that the preference distance is always non-negative, $\overline{D}_{F^i,F^j} \geq 0$ [6].

The average preference distance of the group is then defined as the average of the pairwise preference distances, computed over all $N(N-1)/2$ node pairs in the group:

$$\hat{D}_F = \frac{\sum_{(i,j)} \overline{D}_{F^i,F^j}}{N(N-1)/2} \ ,$$

Finally, we define *tightness* T to be the inverse of the average preference distance of the group:

$$T = \frac{1}{\hat{D}_F}. \quad (2)$$

Tightness expresses the similarity of interests among nodes of the social group and is always greater or equal to zero. $T \to \infty$ when the interests of nodes for the objects coincide, whereas $T \to 0$ implies that the nodes have completely different preferences. As T is an average metric of interest "closeness" among the nodes of a social group, it is clear that a given value of T may arise under different sets of interest distributions of the nodes. In order to draw more insightful conclusions in the current study, we consider the following two cases of dissimilarity in the nodes' interest distributions:

- Case 1: The order (rank) of the objects remains the same for all nodes (i.e., the first-ranked object for all nodes is the same, the second-ranked object is the same, etc). However, the interest distributions are different: they become more concentrated around the most popular objects as the node index n increases.
- Case 2: The interest distributions are identical for all nodes but the ranking of a given object may change for different nodes (e.g., nodes do not necessarily have the same object as their kth-ranked one).

In the numerical examples in this paper $N = 5$ nodes (to better illustrate the results), the capacity $C = 10$ objects, the object population is $M = 50$ objects and $t_l = 0$, $t_r = 10$ and $t_s = 20$ cost units. All preference distributions in the test cases that follow are Zipf distributions. The Zipf distribution has been shown to be a good model for the popularity of web objects [12], which could also constitute the bulk of traffic in our scenario. Moreover, it is remarkably flexible in capturing a wide range of distributions, from the uniform (for skewness parameter $s = 0$) to much more heavy-tail distributions with higher skewness values ($s > 0$).

A. Case 1

The request rates F^n_m of node n over the objects $m = 1, 2, ..., M$ are drawn from a Zipf distribution with different exponent s for each node. We consider that Node 1 has uniform interest distribution, i.e., $s = 0$ for this node. Then, for node n, $n = 2, 3..., N$, s is increased by $p(n-1)$, where $p \in \mathbf{R}$ is the increment parameter. For example, when $p = 0.2$, $s = 0.2$ for Node 2, $s = 0.4$ for Node 3, and so on. The request rate of node n for object m is given by:

$$F^n_m = f(m; s, M) = \frac{1/m^s}{\sum_{l=1}^{M} 1/l^s}. \quad (3)$$

Table I
EXAMPLE TIGHTNESS VALUES

(a) T when increased by different values of p

Increment parameter (p)	Tightness (T)
0.0	∞
0.2	2.0861
0.4	0.4614
0.6	0.2398
0.8	0.1697
1.0	0.1362

(b) T when shifted by different values of k

Shift parameter (k)	Tightness (T)
0	∞
1	0.3688
2	0.2674
3	0.2294
4	0.2089
5	0.1962
6	0.1876
10	0.0012

A preference ranking of the M objects is determined by this distribution – which are accordingly ranked as $[1, 2, ..., M]$ – and is the same for all nodes. As s increases, a node's distribution becomes more concentrated in the first objects. Table I(a) shows the value of *tightness* when the interest distributions are derived as outlined above, for different values of the increment parameter p. Notice that *tightness* decreases as p increases. This is because as p increases, the pairwise distances between any two node distributions increase (the difference in their s parameter is higher).

B. Case 2

The request rates F^n_m are drawn from a Zipf distribution with exponent s with $s = 1$ for all nodes, which is given by (3). If the rank of objects is the same for all nodes, we derive from (2) that $T \to \infty$.

In order to establish dissimilarity in the nodes' interests, we create a different rank of objects for each node. To this end, Node 1 is assigned the rank of objects $[1, 2..., M]$ and this rank is shifted to the left by different positions for each of the other nodes. We consider different values of the shift parameter. In general, the rank of objects of node n, $n = 1, \ldots, N$ is shifted by $k(n-1)$ positions, where $k \in \mathbf{N}$ is the shift parameter. For example, when $k = 1$, the ranking of Node 1 is $[1, 2, ..., M]$, the ranking of Node 2 is $[2, 3, ..., M, 1]$, the ranking of Node 3 is $[3, 4, ..., M, 1, 2]$, and so on. Table I(b) shows the value of *tightness* when the interest distributions are shifted by various positions, as described above. Notice that *tightness* decreases as the shift parameter increases. This is because the average absolute difference of the distributions (for the same object) between any two nodes increases, as k increases. Generally, numerical values of *tightness* are close to zero, and increase abruptly as distributions become more similar.

III. CONTENT PLACEMENT STRATEGIES

In this section we describe the object placement strategies that we consider in this paper. Under the *Optimally altruistic* strategy the objects are stored in such a way that the total access cost for all nodes in the social group is minimized (i.e., minimize $\sum_{n=1}^{N} \mathcal{C}_n(P)$). This problem can be transformed into a 0-1 integer programming problem.

Let $X_m^n = \begin{cases} 1, & \text{if } m \in P_n; \\ 0, & \text{otherwise} \end{cases}$ and

$Y_m^n = \begin{cases} 1, & \text{if } m \notin P_n \text{ and } m \in P_{-n}; \\ 0, & \text{otherwise.} \end{cases}$

The objective is to minimize the function of the total access cost:

$$\sum_{n=1}^{N}\sum_{m=1}^{M} X_m^n F_m^n t_l + Y_m^n F_m^n t_r + (1 - X_m^n)(1 - Y_m^n) F_m^n t_s,$$

where

$$Y_m^n = (1 - X_m^n)(1 - \prod_{\substack{j=1 \\ j \neq n}}^{N} (1 - X_m^j)).$$

This is a quadratic programming problem, whose solution is very difficult. A re-formulation of the above to a linear problem is derived in [9].

Under the *Selfish* strategy, or *greedy local* strategy as referred to [8], the nodes only store their most preferable objects. Each node n ranks the objects in a decreasing order $[1, 2, ..., M]$ such that $F_1^n \geq F_2^n \geq ... \geq F_M^n$ and selects to store the first C_n ones. Thus, $P_n = [1, 2, ..., C_n]$.

Finally, under the *Self-aware cooperative* strategy each node first stores its C_n most preferable objects and then makes replacements based on the placements of the other nodes in order to improve its placement [8]. Thus, a node may decide to evict an object that exists in some other node in the group in order to insert a new object, if this incurs an access cost reduction in (1). As the nodes play sequentially, each replacement made by a node may negatively affect the access cost of other nodes. It is proved in [8] that this strategy is mistreatment-free, i.e., for any node n, it holds that $\mathcal{C}_n^C(P) \leq \mathcal{C}_n^S(P)$, where $\mathcal{C}_n^C(P)$ denotes the access cost of node n for all the objects under the self-aware cooperative strategy and $\mathcal{C}_n^S(P)$ denotes its access cost under the selfish one. Thus, the final access cost of a node is at most as high as its access cost under the selfish strategy.

It was shown in [8] that only objects not already included in the group can be inserted during a replacement step, evicting only objects which are present elsewhere in the group (duplicates). Due to these properties (and as the results followed have shown) the performance of the self-aware cooperative strategy can be very close to the optimally (social-cost minimizing) altruistic one. In Section IV we explore under which conditions this performance is achieved.

On a more practical note, an optimally altruistic behavior requires complete knowledge of the group's characteristics (demand patterns of all nodes in the group) and a solution to the global optimization problem. This can be solved by a central authority that dictates its placement decisions to the nodes, or in a distributed fashion, in which each node solves the global optimization problem and then all nodes negotiate their placements (there are multiple solutions to the global optimization problem). The self-aware cooperative strategy requires less information; more implementation details can be found in [8]. The selfish strategy has the smallest implementation cost, since each node does not need to be aware of the group's interests.

IV. Numerical Evaluation

In this section we present some numerical examples to illustrate the impact of *tightness* on the access cost under the three behavior-based content placement strategies. The conclusions drawn from these results can help establish clear guidelines as to which strategy would be beneficial to the individual nodes and/or the entire group, and under which tightness conditions in the group.

Fig. 1 and 2 show the individual node and total (i.e., for the entire group) access cost under the different content placement strategies and for different values of *tightness*. Both scenarios for the interest distribution dissimilarity are considered (Case 1 and Case 2, Section II).

A. *Social groups with infinite or very high* tightness

The results show that the optimally altruistic strategy is the best performing one regarding both the individual cost for any node (Fig. 1(a)), as well as the cost for the entire group (Fig. 2, $T = \infty$), for both Cases 1 and 2. Consequently, the optimally altruistic behavior is the clear winner-behavior for any node in a very tight social group.

Notice that under very high *tightness*, the individual and total access cost induced by the self-aware cooperative strategy is (a) very close to (slightly higher than) that under the optimally altruistic and (b) is always lower than that under the selfish strategy. In other words, the self-aware cooperative strategy induces no node mistreatment while yielding performance close to the optimal. Thus, given its lower implementation complexity compared to the optimally altruistic (see Section III), it may be selected as an easier to implement and similarly performing alternative to the optimally altruistic strategy in very tight social groups.

To this end, the larger the *tightness* of the social group, the greater the group's benefits when nodes are either cooperative or optimally altruistic, compared to being selfish.

B. *Social groups with very low* tightness

While it was shown that under very high *tightness* both the individual nodes and the entire group will benefit by having the nodes adopt the optimally altruistic or the self-aware cooperative strategies (compared to the selfish one), Fig. 1(d) and 2 (for $T \simeq 0$) show that this is not the case under low or very low *tightness* of the social group.

Although the optimally altruistic strategy (always) brings the maximum benefits for the group, it mistreats individual nodes under such group tightness conditions. (e.g., Node 5 in Fig. 1(d), for both Cases 1 and 2). Furthermore, the benefits to the group are about the same as under the selfish strategy or

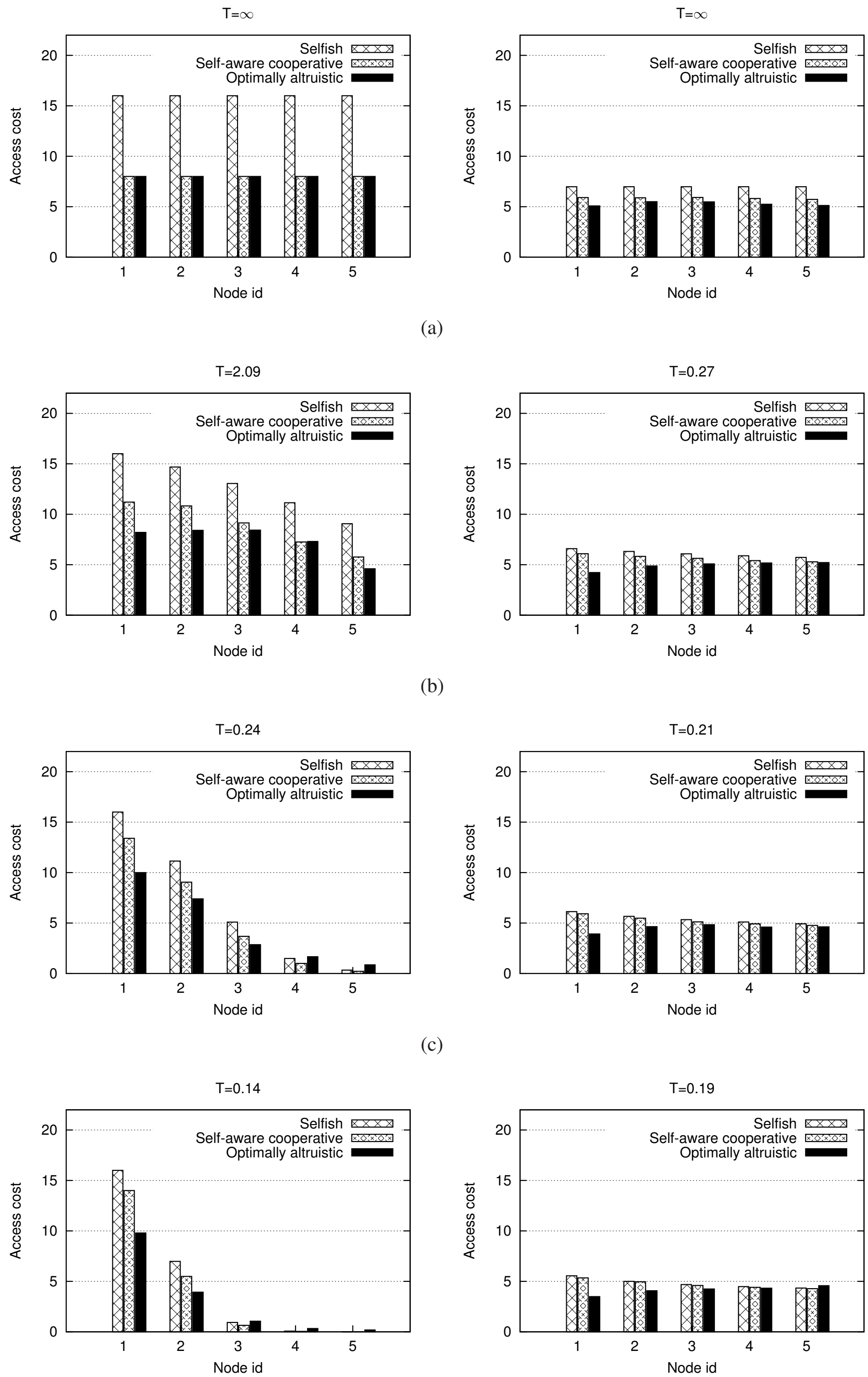
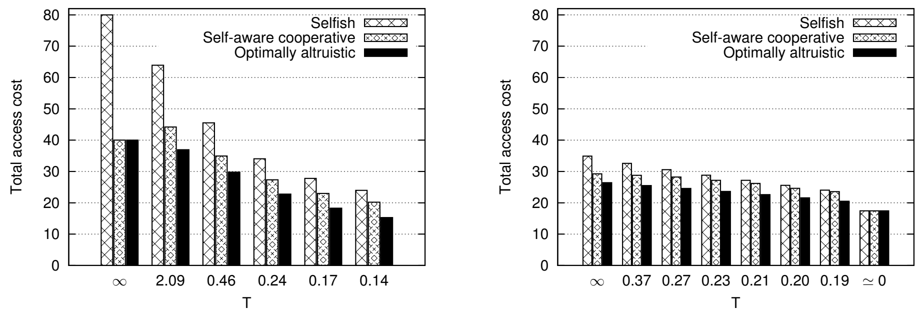


Figure 1. Individual access cost under different strategies for different values of *tightness* T, under Case 1 (figures on the left) and Case 2 (figures on the right)



(a) (b)

Figure 2. Total access cost under different strategies for different values of *tightness* T, under Case 1 (figure on the left) and 2 (figure on the right)

only slightly greater (Fig. 2(b), $T \simeq 0$ or Fig. 2(a), $T = 0.14$). Since a) the implementation of the optimally altruistic strategy is considerably more complex than the selfish one (see Section III); b) it does not avoid mistreating certain nodes, threatening the stability of the social group; and c) it only brings small benefit to the group compared to the selfish one, it may be concluded that the selfish strategy should be preferred against the optimally altruistic.

Finally, it should be noted that both the self-aware cooperative and selfish strategies are mistreatment-free. From Fig. 2 it is clear that the self-aware cooperative strategy will bring negligible benefits if any (the same (Fig. 2(b), $T \simeq 0$) or only slightly larger (Fig. 2(a), $T = 0.14$)) to the social group compared to the selfish one. Thus, given that the complexity of the self-aware cooperative strategy, requiring information exchange among the nodes, is substantially larger than that under the selfish one, requiring no information exchange and lower computational complexity, it may be concluded that, under low *tightness*, the selfish strategy should be preferred against the self-aware cooperative.

V. Conclusion and Future Work

In this paper we investigated how the commonality in the social interests of nodes in a social group affect the performance of content storage strategies. Three such strategies capturing a broad spectrum of behavior-shaped content storage were considered: the selfish, self-aware cooperative and optimally altruistic one. We first proposed a new metric for measuring the commonality in interests (social similarity), called *tightness*, based on the mean value of the Kullback-Leibler divergence of nodes' preferences. We then varied tightness across its value range and compared the three content storage strategies, with respect to the individual and total access costs they achieve.

Given the coherence of the obtained results for two different cases of dissimilarity in the interest distributions (Case 1 and 2), it appears that *tightness* is an effective metric for deciding which behavior (or strategy) a node should adopt.

More specifically, altruism is a win-win virtue or behavior only in tight social groups (if the implementation cost is not an issue): both the group's benefit and individual's benefits exceed those under a self-aware cooperative or a selfish behavior. As *tightness* decreases, smaller group benefits can be induced either through altruism or a self-aware cooperative behavior, while mistreatment may incur under altruism. Thus, in this case, acting selfishly ensures no mistreatment and no significant loss to the group's benefit.

A possible line of future work is to collect user preference data from a real social network, in order to fit these more accurately to Zipf-like distributions. Further, we plan to study the impact of group size on the global and individual access cost, as a function of the preference distributions.

References

[1] C. Boldrini, M. Conti, and A. Passarella. Contentplace: Social-aware data dissemination in opportunistic networks. In *The 11-th ACM (MSWiM'08)*, 2008.

[2] I. Carreras, D. Tacconi, and A. Bassoli. *Social Opportunistic Computing: Design for Autonomic User-Centric Systems*, pages 211–+. 2009.

[3] M. Conti and M. Kumar. Opportunities in opportunistic computing. *Computer*, 99(PrePrints), 2009.

[4] L. Hossain, K. S. K. Chung, and S. T. H. Murshed. Exploring temporal communication through social networks. In *INTERACT (1)*, pages 19–30, 2007.

[5] E. Jaho and I. Stavrakakis. Joint imterest- and locality-aware content dissemination in social networks. In *Sixth Annual Conference on IFIP/IEEE WONS*, 2009.

[6] S. Kullback. *Information theory and statistics*. John Wiley and Sons., 1959.

[7] K.-W. Kwong, A. Chaintreau, and R. Guerin. Quantifying content consistency improvements through opportunistic contacts. In *CHANTS '09*, pages 43–50, New York, NY, USA, 2009. ACM.

[8] N. Laoutaris, O. Telelis, V. Zissimopoulos, and I. Stavrakakis. Distributed selfish replication. *IEEE Trans. Par. Distr. Systems*, 17(12):1401–1413, Dec. 2006.

[9] A. Leff, J. Wolff, and P. Yu. Replication algorithms in a remote caching architecture. *IEEE Trans. Par. and Distr. Systems*, 4(11):1185–1204, Nov.

[10] J. L. Myers and A. D. Well. *Research Design and Statistical Analysis (2nd edition)*. Lawrence Erlbaum, November 2003.

[11] J. P. Scott. *Social Network Analysis: A Handbook*. SAGE Publications, January 2000.

[12] D. N. Serpanos, G. Karakostas, and W. H. Wolf. Effective caching of web objects using zipf's law. In *IEEE International Conference on Multimedia and Expo (II)*, pages 727–730, 2000.

[13] J. Vegelius, S. Janson, and F. Johansson. Measures of similarity between distributions. *Quality and Quantity*, 20(4):437–441, 1986.

[14] K. Xu, P. Hui, V. O. K.Li, J. Crowcroft, V. Latora, and P. Lio. Impact of altruism on opportunistic communications. In *ICUFN'09*, pages 153–158, 2009.