Social Communication to Improve Group Recognition in Mobile Networks

Anna V. Guglielmi and Leonardo Badia
Dept. of Information Engineering, University of Padova, via Gradenigo 6B, 35131 Padova, Italy
email: {guglielm, badia}@dei.unipd.it

***Abstract*—In this paper, we consider the problem of recognizing groups of similar nodes within a mobile network. Nodes belonging to the same group have the same preferences, e.g., in terms of friendship with other nodes. However, due to limited communication range, each node has a partial view of the network and is aware only of its friendship relationships with the immediate neighbors. Depending on how large the communication range actually is, the problem can be in principle solved by clustering nodes, classifying the friendship relationships on a scale from -1 to $+1$ (sure not friendship to sure friendship, with unknown nodes classified as 0). However, if the communication range is small, this approach leads to large errors. We show how a simpler procedure based on the exchange of social information in a collaborative fashion is able to achieve better results than more expensive clustering algorithms. Also, the two approaches can be combined by clustering nodes after exchanging a given amount of social information, thus improving the overall results.**

***Index Terms*—Social network analysis, clustering, profiling, graph theory, Internet of Everything.**

I. Introduction

ONLINE social networks are still enjoying an unprecedented success and represent a very effective instrument for mass communication and content dissemination over the Internet. From the scientific research standpoint, a strong theoretical component may be envisioned behind the pattern recognition and profiling engines that many online social networks adopt to offer better services to their users [1].

Profiling users can also be useful to discover communities of individuals sharing some attributes and therefore exploiting the inherent redundancy within the network, for example, to provide them with similar services or caching homogeneous contents [2], [3]. However, profiling procedures also pose several challenges about anonymity and privacy of the users [4], [5]. In particular, among the many challenges related to this problem, we focus on the detection of group structures, which is an interesting application of social network analysis. From the mathematical standpoint, the identification of group structures in social networks can be seen as a clustering optimization problem [6]–[8]. Also in the context of mobile networking, clustering has been widely investigated and is not an entirely new problem [9], [10]. However, what makes our investigation for social networks challenging is that we do not want to cluster mobile nodes based on their technical attributes (such as radio technologies or energy consumption) but rather on some independent application-driven parameters, such as their individual preferences or the societal relationship among them (e.g., who are their "friend nodes").

Most of the theoretical analysis in this field is implicitly based on the assumption that online social networks are able to collect massive amounts of information on their users. The problem resides in mining these data, rather than improving their availability, since the network owner is generally able to collect lots of data on every single user; most of the times, the users themselves are willingly sharing their preferences and attributes across the entire network [11].

In this paper, we adopt a different perspective by considering a mobile network where discovery of these individual traits of the users is only partial. There are several possible motivations for this scenario; we argue that mobility can be a primal cause. For example, it is reasonable to assume that mobility hinders communication and therefore limits the collection of data. At the same time, mobile networks can also have a temporary character and therefore most of the nodes only gain knowledge on their immediate neighbors only.

We remark that the two attributes of "mobile" and "social" networking have a complementary relationship, in that for example the latter can improve the former. In some papers [7], [12] it is studied how social relationships can help the routing algorithms. Again, this approach can be seen as a byproduct of context-aware networking. At the same time, social connections are used, for example by [2], as a way of further profiling users, since it can be assumed that users are likely to connect with each other if they share similar preferences, and their behavior is akin to those who they recognize as friends.

The novel aspect of this paper is that we see mobility of the nodes as a *detrimental* to the information extraction, since it causes the network view to be only partial. In this sense, our claim is that the problems caused by such a partial information can only be overcome thanks to social collaboration among the nodes. Thus, we investigate how the collaborative exchange of information about each user's preferences can help recognizing the group structures in the network.

In particular, the scenario on which we focus, and that will be better detailed in the following, involves a network where nodes are element of a graph (representing a network snapshot) and are classified a priori as belonging to a group, which, in graph theory terms, is seen as a *color*. Colors have pre-defined bindings, which can be regarded as a binary friendship relationship over the social networks. However, nodes only have awareness of their relationship with their

immediate neighbors on the graph. Since the network is not fully connected, the extraction of group information is therefore noisy. Here is where social collaboration applies. We devise a simple iterative procedure where the estimate about unknown nodes is further improved through periodic reporting that nodes make about their neighbors, on a random basis, as a social diffusion of context information.

We believe that such a procedure can be applied in any scenario where social exchange of information is involved, but a centralized coordination is difficult and therefore it is more convenient to resort to distributed exchange of signaling among the nodes. This may be the case for machine-to-machine communication in vehicular scenarios [13], that can be applied to exchanging road safety information, but where exchanges are only made among devices with compatible service (which represents our friendship relationship). For the Internet of Everything, disconnections can be experienced but social relationships among the nodes can be exploited to establish a routing backbone [14].

As a result, we show that, while estimating clusters from the initial incomplete data can be highly ineffective, the iterative procedure is much more accurate in identifying clusters and classifying individual nodes' preferences. A combination of this procedure used as a preliminary scan to improve the data, and a standard clustering approach applied on top, is even more effective. Since our goal is to show the effectiveness of social collaboration as a principle, both the diffusion procedure and the clustering technique are intentionally kept simple; their detailed analysis would be out of the scope of the present paper. However, both strategies can actually be refined and different choices of the pattern recognition technique can be made [8]. From our preliminary investigations we expect the effectiveness of social collaborative networking to hold with more complex strategies.

The rest of this paper is organized as follows. In Section II we formalize the model under investigation. In Section III we describe our proposed approach, as opposed to a standard k-means clustering, to recognize groups of nodes with similar preferences. Section IV shows the numerical results. Finally, Section V concludes the paper.

II. System Model

We consider a network with N nodes spread over a planar area, that are labeled with integer numbers $1, 2, \ldots, N$. For simplicity, we consider that neighborhood relationships of the network nodes are based on a distance criterion. In particular, we define a coverage radius r and we define two nodes to be neighbors iff their Euclidean distance is lower than r. Note what follows. First of all, it is not restrictive to consider a higher number of dimensions, or a different distance than the Euclidean one. Also, our distance-based criterion can be directly linked with more realistic connectivity relationships based on path-loss and signal attenuation, which ultimately depend on the distance between the transmitter and the receiver. Naturally, more realistic radio communication models [15] would also include other effects such as shadowing and fading, and capture node interference; still, the result will be analogous. In other words, we do not claim that a simple unit-disc model is particularly suitable for wireless communication [16]; it is just that the specific investigations made in this paper do not critically depend on this characterization, which, however, offers the advantage of representing network connectivity with a simple parameter, the coverage radius r.

The network connectivity as a graph can be represented by an $N \times N$ adjacency matrix $\mathbf{A}$, where entry a_{ij} at the ith row and jth column of $\mathbf{A}$ is equal to 1 if i and j are neighbors (specifically, they are less than r apart) and 0 otherwise.

Furthermore, we assume the existence (a priori) of C non-overlapping groups of nodes, which randomly partition the set of the N nodes. For the ease of terminology, we define each of the C groups as a "cluster" or also a "color" (we will use these two terms interchangeably), so that we can see the recognition of group structure as a clustering problem or a graph coloring problem. The color of each node is assumed to be randomly and independently determined. In the following, we will also assume that all the colors are equiprobable for a node, even though this is not necessary for our procedures. The color of each node can be described by vector $\mathbf{v}$ where ith element v_i is an integer between 1 and C which is the id of i's cluster.

Colors are also associated with relationship, which will be referred to as "friendship." Note that this naming is just again chosen for the sake of exposition, it can actually refer to any reflexive and symmetric (but not necessarily transitive) property that can relate a pair of nodes. In particular, we assume that a node is a friend of itself, and all nodes of the same colors are friends. Also, the friendship relationship (or lack thereof) is symmetric. However, being friends is just a necessary (but not sufficient) condition to belong to the same cluster; also, friendship is not transitive since we allow for any pair of nodes with respective colors a and b to be friends with each other, and the same if their colors are b and c, without necessarily implying that any two nodes of respective colors a and c are also friends. The specific aspect of our paper is that, due to mobility, this relationship among an arbitrary pair of nodes is not known in general; it is so only for neighbor nodes, and the goal of our algorithms is to infer this relationship correctly also between non-neighbors [2], [11].

Importantly, the naming of this relationship just mimics an important relationship on online social communities (alternatively dubbed as belonging to the same "circle" or "professional network") but, at least to the extent of the present paper, has nothing to do with collaborative behavior among the nodes, which we assume to exist regardless of their friendship relationship; indeed, we will implicitly show that it is beneficial also for non-friend nodes to collaborate. As a possible extension of the present paper, it is possible to consider a scenario where social collaboration is limited to actual friend nodes. However, our connection of same-color nodes as belonging to the same "logical" group is entirely separate from (and actually independent of) considering "physical" relationships of neighborhood [12].

We can define a $C \times C$ symmetric matrix $\mathbf{B}$ to describe the relationships between different clusters, in particular its entry b_{fg}, with $f, g \in \{1, 2, .., C\}$, is set to $+1$ if the nodes in cluster f and those in cluster g are friends, and to -1 otherwise. Note

that the choice of the two values representing friends and non-friends (in the following also called "enemies") is arbitrary, but using $+1$ and -1 allows for a consistent mathematical representation of the remaining procedures. In particular, value 0 will represent indeterminacy between the two values.

We further remark that, to have an interesting problem, we require that all the rows in matrix $\mathbf{B}$ are different, since otherwise there were two colors that have identical relationship with the others, and therefore can be merged in a single one. Also, due to the reflexive requirement, matrix $\mathbf{B}$ must have all diagonal elements equal to 1. However, these properties still leave open the characterization of $\mathbf{B}$ for the off-diagonal elements. Intuitively speaking, a case where $\mathbf{B}$ contains many elements equal to -1 is easier to manage, since it means that nodes are only friends with those of their same color and therefore knowing that a neighbor node is a friend is more descriptive. In the following, we will consider two different choices to determine $\mathbf{B}$. We will always consider it to be an arbitrarily determined full-rank symmetric matrix with an all-1 diagonal, but we constrain it to have either 25% or 50% of its elements equal to 1. We refer to these cases as "low inter-color friendship" and "high inter-color friendship," respectively.

Generalizing the idea behind matrix $\mathbf{B}$, we can also define an $N \times N$ matrix $\mathbf{F}$, which describes friendship between each node of the network. Specifically, we can combine the information in $\mathbf{v}$ and $\mathbf{B}$ to derive every entry in $\mathbf{F}$; for example, we can notice that

$$f_{ij} = b_{v_i v_j}. \tag{1}$$

However, as clear from the premises, such a matrix $\mathbf{F}$ is not available. We only know a partial estimate $\widetilde{\mathbf{F}}$, where non neighbor nodes do not have a 1 or -1 at their entry. Formally,

$$\widetilde{\mathbf{F}} = \mathbf{F} \odot \mathbf{A} \tag{2}$$

where $\odot$ denotes the element-wise scalar multiplication.

III. Proposed Social Communication Technique

The purpose of the present paper is to identify, based only on $\widetilde{\mathbf{F}}$, an estimate $\widehat{\mathbf{F}}$ of the full-network matrix $\mathbf{F}$. Since the real $\mathbf{F}$ is not used but is actually known, it can be compared with its estimate $\widehat{\mathbf{F}}$ by checking the Frobenius norm $||\widehat{\mathbf{F}}-\mathbf{F}||_{\mathrm{F}}$. To have a normalized value and also since we can remark that the accuracy of the diagonal elements is always 100%, we define the following estimate efficiency parameter η as

$$\eta = 1 - \frac{||\widehat{\mathbf{F}} - \mathbf{F}||_{\mathrm{F}}}{2N(N-1)}, \tag{3}$$

so that maximum efficiency of 1 is reached when all the elements of the estimate $\widehat{\mathbf{F}}$ coincide with those of $\mathbf{F}$.

A straightforward solution to our problem is to apply an out-of-the-box clustering procedure. Indeed, the elements of $\widetilde{\mathbf{F}}$ can be seen as points in an N-dimensional space. In this sense, indeterminate points have been set to 0, which is the intermediate value between possibilities $+1$ and -1; thus, a direct application of a clustering process makes sense. For example, we can apply a k-means clustering [9], which is a standard reference technique for this kind of investigation. As a side note this choice poses some theoretical questions, especially in that it requires to know in advance the number of clusters. However, we choose it due to its generality and widespread usage as a comparison term. It also involves, as many other clustering techniques, a direct drawback if applied to this scenario. In fact, applying such a clustering scheme would totally ignore the underlying social structure of the network, treating each row of $\widetilde{\mathbf{F}}$ as a different point in the data space. We expect that the conclusions drawn for the k-means technique would be applicable also to different and possibly more sophisticated clustering procedures, such as the Chinese restaurant clustering or affinity propagation [8], [17]. The clustering procedure does not directly determine $\widehat{\mathbf{F}}$, which would be needed to apply (3), but rather gives a cluster id for all the nodes, which can be seen as an estimate $\widehat{\mathbf{v}}$ of vector $\mathbf{v}$, and also an estimate of $\mathbf{B}$ by considering the centroids of the k-means. These points do not generally have values equal to either -1 or $+1$, but can also have intermediate values. Thus, it is required to round them (up or down, accordingly) to the closer one. Finally, $\widehat{\mathbf{F}}$ is derived, see (1), as

$$\widehat{f}_{ij} = \widehat{b}_{\hat{v}_i \hat{v}_j}. \tag{4}$$

It will be shown in the next section that the k-means clustering does not achieve a very good performance. Thus, our goal is to see whether we can improve the efficiency of group detection by considering a system in which the nodes can communicate with each other and update their current friendship estimates based on the information that the other nodes spread in the network. To this end, we develop an iterative algorithm where we set an initial matrix $\mathbf{X}^{(0)} = \widetilde{\mathbf{F}}$ and we update its entries from the information communicated by a row (representing a node disseminating its data) chosen at random, so as to obtain matrix $\mathbf{X}^{(1)}$. After that, another row of $\mathbf{X}^{(1)}$ is chosen and the iteration is repeated to obtain $\mathbf{X}^{(2)}$ and so on. In principle, this can keep going on until a matrix $\mathbf{X}^{(T)}$ is obtained where all the elements are either $+1$ or -1. However, the evaluation of the efficiency is based on the Frobenius norm that works element-wise and can be computed even if some elements have not reached the border values. We find out that it is more convenient for computational reasons to perform a given number of iterations; generally, we can get very close to the stopping point of the algorithm even in this way, but with a faster execution time.

We propose a policy based on a social collaborative exchange of information, where we also involve the following parameters: a threshold k chosen between 0 and 1, an integer value M, which jointly determine whether to update a value or not, and finally a real positive value Δ that describes the (initial) amount of the update.

Each iteration works in the following way. We choose at random the pth row of the temporary estimate $\mathbf{X}^{(t)}$ and we update from it all other rows (i.e., the friendship estimate of all other nodes) to determine $\mathbf{X}^{(t+1)}$. By looking at $\mathbf{X}^{(t)}$, we take a generic row i and we call those nodes j for which element $\widehat{f}_{ij} = +1$ as the *sure friends* of i, and similarly we define those for which $\widehat{f}_{ij} = -1$ as its *sure enemies*. Threshold k is used to label those nodes j that at the ith row have an entry $\widehat{f}_{ij} > +k$ as *likely friends* and similarly those for which $\widehat{f}_{ij} < -k$ are said to be *likely enemies*.

Note that these are just labels used while the algorithm is running. At the beginning of the iterations, all sure friends or enemies are identified correctly since they are based on authentic values of $\mathbf{F}$. After some iterations, there is no guarantee that this labeling is correct anymore.

To decide whether a specific row q different from p can be updated, we consider the element-wise product between the pth and qth rows. The result of this product can give three different outcomes: *o1)* node q likely has the same color of node p; *o2)* node q and node p likely belong to different clusters; *o3)* there is nothing that can be said. To call the first outcome, *o1*, we must have all the sure friends of p (i.e., their corresponding entry of the pth row is equal to $+1$) are not that sure enemies of q, and also the converse relationship must hold (sure enemies of p cannot be sure friends of q). Moreover, node q cannot have a likely friend that is a sure enemy for node p, or node q cannot have a likely enemy that is a sure friend for p. If any of these conditions is violated, then the outcome is *o2*. However, we do not consider sufficient not to violate any of these condition to have outcome *o1*; for example, the element-wise product can be entirely made by zeros (if node p and q do not share any neighbors) which is of little meaning but does not violate any constraint.

In our proposed procedure, *o2* is declared as the outcome if at least one vector element is negative. Else, *o1* is declared if at least M elements of the product vector are greater than k. Otherwise, the outcome is declared to be *o3*. Only if *o1* is declared, an update is performed for the pth row of the matrix, so that every qth element, for $q \in \{1, 2, .., N\}$ whose value $\mathbf{X}_{pq}^{(t)}$ is strictly between -1 and $+1$ is added the following quantity: $\Delta \cdot \mathbf{X}_{pq}^{(t)}$. As can be observed, being $\Delta \in (0, 1)$ this operation is actually an addition if nodes p and q are friends, it is a subtraction otherwise. Also, if the updated value exceeds the absolute value of 1, it is capped to either -1 or $+1$.

The procedure can be iterated until a matrix $\mathbf{X}^{(t)}$ is obtained with only $-1/+1$ values, or when $t = T_{\max}$. Then, we set $\widehat{\mathbf{F}} = \mathbf{X}^{(t)}$. Algorithm 1 shows the pseudocode considered for our evaluations, where the algorithm ends after $T_{\max}$ updates of the matrix. Note that the input value denoted as $X(:, :, 0)$ corresponds to $\mathbf{X}^{(0)} = \widetilde{\mathbf{F}}$ and is derived from the network topology and the real nodes' colors, in particular from parameters N, r, $\mathbf{A}$, $\mathbf{B}$ as per the aforementioned unit disc model and equations (1) and (2). The output value $X(:, :, T_{\max})$ in the pseudocode corresponds therefore to $\widehat{\mathbf{F}} = \mathbf{X}^{(T_{\max})}$.

Another possibility of applying our rationale towards an improved estimate than that obtained by clustering the raw data is to still apply a clustering algorithm, in particular we use again the k-means procedure, but after an initial phase where social communication has improved the data. In other words, we run our proposed heuristic iterative procedure for a given number of times.

We update matrix $\mathbf{X}^{(t)}$ but (regardless of whether convergence is perfectly reached or not) we apply at the end a k-means clustering algorithm. Our goal with this last strategy is to verify whether the iterative procedure of social collaboration improves the data per se, and enables the clustering algorithm to obtain a more efficient result.

Algorithm 1 Pseudocode of the iterative process

```
Input : N, M, k, Δ, T_max, X(:, :, 0)
Output : X(:, :, T_max)
t ← 0
while t ≤ T_max do
    randomly choose n ∈ {1, 2, ..., N}
    for every m ∈ {1, 2, ..., N} \ {n} do
        a(:) = X(n, :, t) ⊙ X(m, :, t)
        ℓ ← 0
        for i ∈ {1, 2, ..., N} do
            if a(i) > k then
                ℓ ← ℓ + 1
            end if
            if a(i) < 0 then
                ℓ ← 0
                break
            end if
        end for
        if ℓ ≥ M then
            for every node i do
                X(n', i, t) = X(n', i, t) + Δ · X(n, i, t)
                if X(n', i, t) > 1 then
                    X(n', i, t) = 1
                end if
                if X(n', i, t) < −1 then
                    X(n', i, t) = −1
                end if
            end for
        end if
    end for
    t ← t + 1
end while
```

IV. Performance Analysis

We present some numerical results, all obtained for a scenario with a square area of side equal to 100 reference units. In particular, our purpose is to compare the performance considering a system that exploits the collaborative spreading of information in the network and another that does not, and just uses the clustering on the initial data.

We set the parameters introduced in the previous sections as $k = 0.4$, $M = 4$, and $\Delta = 0.01$. These values were chosen by heuristic trial-and-error procedures. Moreover, for this comparison we assume different values for the coverage radius r, the number C of clusters, and the number N of nodes in the entire networks.

All the figures shown consider the normalized efficiency parameter defined in (3) and compare three different approaches: the plain k-means algorithm; the heuristic iterative procedure based on collaborative exchange of information about friendship between nodes; finally, a joint approach where the iterative procedure is preliminary applied, and afterwards a k-means clustering is employed. Every result has been derived from a very large number of simulation runs, so that confidence intervals of the results, not shown in the plots, are extremely narrow.

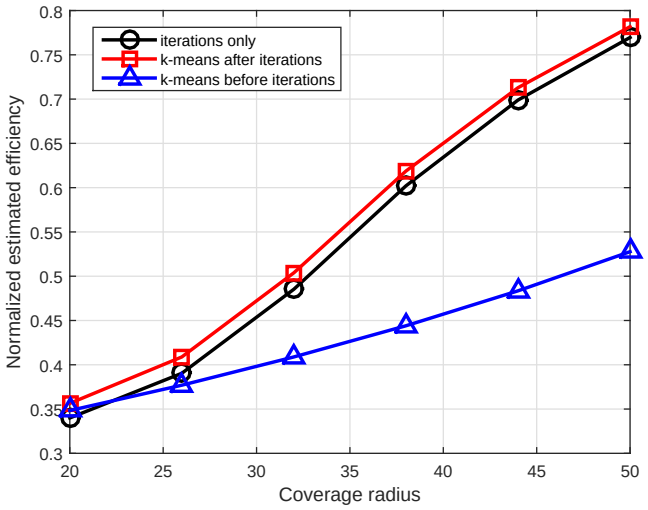

Fig. 1. Normalized classification efficiency as a function of the coverage radius for $N = 200$ nodes, $C = 8$ colors, high inter-color friendship.

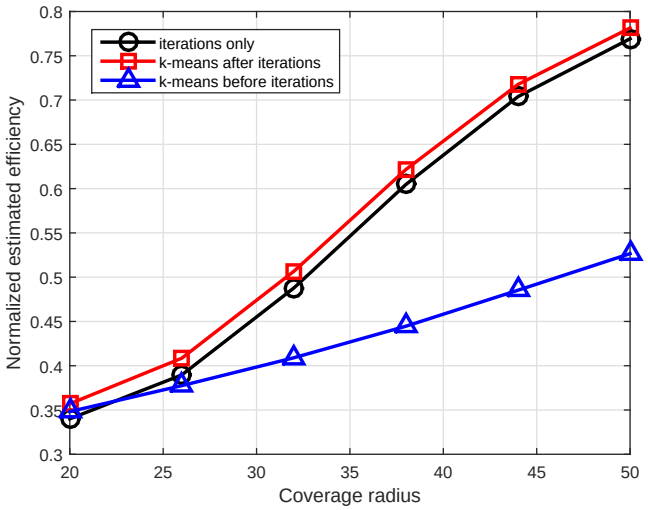

Fig. 2. Normalized classification efficiency as a function of the coverage radius for $N = 200$ nodes, $C = 8$ colors, low inter-color friendship.

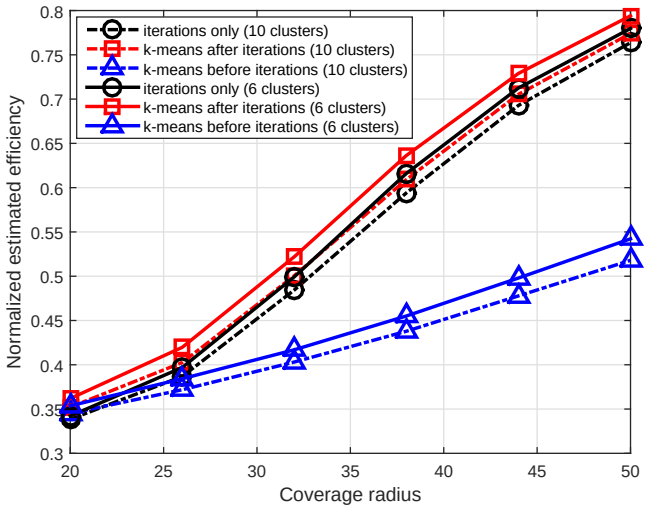

Fig. 3. Normalized classification efficiency as a function of the coverage radius for $N = 200$ nodes, $C = 6; 10$ colors, high inter-color friendship.

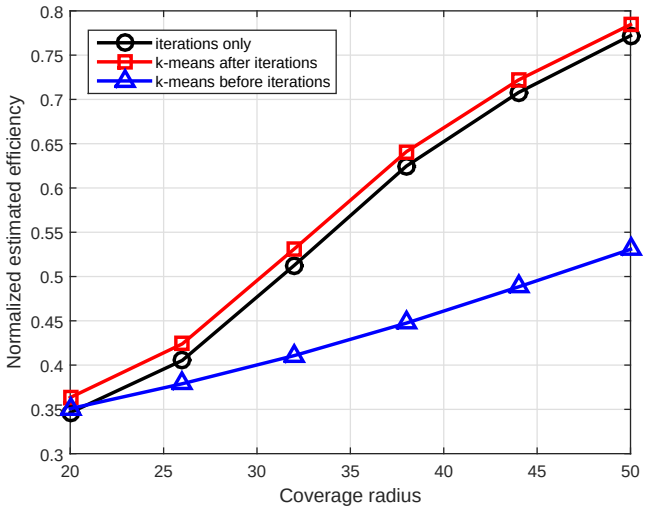

Fig. 4. Normalized classification efficiency as a function of the coverage radius for $N = 500$ nodes, $C = 8$ colors, high inter-color friendship.

Figs. 1 and 2 show the results obtained considering a network with 200 nodes, 8 clusters, and high and low inter-color friendship respectively. A general trend, found in all the results, is that an increase of the communication range r, which causes matrix $\mathbf{A}$ to become less sparse and therefore the initial estimate $\widetilde{\mathbf{F}}$ to be more similar to $\mathbf{F}$, generally improves the efficiency of the estimate; this conclusion holds true for all the three compared approaches, albeit to a different extent.

From the results, it is clearly visible that our proposed approach outperforms the standard k-clustering by increasing the estimate efficiency. The application of the k-means algorithm in order to update again the estimate of $\mathbf{F}$ after the iterations improves the efficiency even more, albeit only slightly. Also, even though the differences are not that striking, the case with high inter-color friendship is characterized by a more difficult recognition of the clusters, especially for the k-means algorithm. In the following, we will report only results related to this case; we checked also the cases with low inter-color friendship and similar conclusions can be drawn there as well.

Fig. 3 allows the comparison of the results obtained changing only the number of clusters considered. Therefore, we consider not only the case with 8 clusters, but also the cases with 6 and 10 clusters. As it can be noted, considering a different number of clusters changes the corresponding efficiency value for each coverage radius. In particular, we can state that a higher number of clusters involves a lower efficiency, and therefore a higher error.

Finally, Fig. 4 compares the results obtained by changing the number of nodes from 200 to 500 in the network. Comparing Figs. 1 and 4, we infer that increasing the number of nodes does not significantly change the performance. Therefore, we can conclude that these algorithms are scalable with respect to the number of nodes in the network, and can be applied even to large scale scenarios of the Internet of Everything.

V. Conclusions and Future Developments

We considered a problem of group detection in social mobile networks, where network mapping by each individual node is assumed to be partial, i.e., limited to its neighbors. We showed how traditional clustering procedures fail to efficiently identify the existing group structures, while a simple heuristic procedure based on the concept of social diffusion of information achieves a much higher efficiency of cluster recognition. Also, the efficiency is further enhanced by combining our proposed procedure with a standard clustering technique.

The work reports preliminary results related to a standard k-means clustering technique and a simple heuristic approach, since the purpose is to show the benefit of social communication. More results can be shown also exploring the parametric dependence of the inner values of the proposed technique (e.g., the update step Δ or threshold k). Future developments also include the evaluations of more sophisticated clustering techniques, especially avoiding the need of k-means clustering to know the number of colors in advance (an element that is, however, not required by our proposed original approach).

Moreover, different heuristic approaches can surely be investigated. While keeping the same rationale of a social communication where nodes collaboratively exchange information about their relationship so as to inform distant nodes, more complex techniques can be thought of, especially exploiting the inherent characteristics of the assumed friendship relationship. For example, contradictions in the self-declared friendship (node i reports being a friend of node j while node j reports the opposite about node i) should be identified and resolved.

Finally, in this paper the logical relationship of collaboration is (intentionally) kept separate from routing aspects and friendship connections. However, one can think of applying the same rationale to scenarios where these aspects are interconnected, i.e., for example nodes only forward data coming from their friends, or they may even disseminate false or malicious information about their enemies. All these open challenges demonstrate the importance of this kind of quantitative investigations to realize efficient paradigms for the Internet of Everything.

References

[1] J. Golbeck, "Trust and nuanced profile similarity in online social networks," *ACM Transactions on the Web*, vol. 3, no. 4, art. 12, Sept. 2009.

[2] R. Li, C. Wang, and K. C.-C. Chang, "User profiling in an ego network: co-profiling attributes and relationships," *Proceedings ACM WWW*, pp. 819–830, 2014.

[3] Z. Wang, D. Zhang, X. Zhou, D. Yang, Z. Yu, and Z. Yu, "Discovering and profiling overlapping communities in location-based social networks," *IEEE Trans. Syst., Man, and Cyb.*, vol. 44, no. 4, pp. 499–509, Apr. 2014.

[4] C. Diaz, C. Troncoso, and A. Serjantov, "On the impact of social network profiling on anonymity," *Lect. Notes in Comput. Science*, vol. 5134, pp. 44-62, 2008.

[5] L.A. Cutillo, R. Molva, and T. Strufe, "Safebook: A privacy-preserving online social network leveraging on real-life trust," *IEEE Commun. Mag.,* vol. 47, no. 12, pp. 94–101, Dec. 2009.

[6] Q. Cai, M. Gong, L. Ma, S. Ruan, F. Yuan, and L. Jiao, "Greedy discrete particle swarm optimization for large-scale social network clustering," *Information Sciences*, vol. 316, pp. 503–516, Sept. 2015.

[7] S. Sharma and S. K. Jena, "Cluster based multipath routing protocol for wireless sensor networks," *ACM SIGCOMM Comp. Commun. Rev.*, vol. 45, no. 2, pp. 14–20, Apr. 2015.

[8] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 97217976, 2007.

[9] P. Sasikumar and S. Khara, "K-means clustering in wireless sensor networks," *Proc. Int. Conf. Comput. Intellig. Commun. Netw.*, pp. 14017144, Nov. 2012.

[10] M. Rossi, L. Badia, P. Giacon, M. Zorzi, "Energy and connectivity performance of routing groups in multi-radio multi-hop networks," *Wirel. Commun. Mob. Comput.*, vol. 8, no. 3, pp. 327-342, Mar. 2008.

[11] L. Gou, M. X. Zhou, H. Yang, "KnowMe and ShareMe: understanding automatically discovered personality traits from social media and user sharing preferences," *Proc. ACM CHI*, pp. 955–964, 2014.

[12] M. Xiao, J. Wu, L. Huang, "Community-aware opportunistic routing in mobile social networks," *IEEE Trans. Comp.*, vol. 63, no. 7, pp. 1682–1695, July 2014.

[13] I. Lequerica, M. García Longaron, P. M. Ruiz, "Drive and share: efficient provisioning of social networks in vehicular scenarios," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 90-97, Nov. 2010.

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

[15] L. Badia, A. Erta, L. Lenzini, M. Zorzi, A general interference-aware framework for joint routing and link scheduling in wireless mesh networks," *IEEE Network*, vol. 22, no. 1, pp. 32-38, Jan. 2008.

[16] M. M. Halldórsson, "Modeling reality algorithmically: the case of wireless communication," *Lect. Notes in Comput. Science*, vol. 8243, pp. 1-5, Dec. 2013.

[17] C. Jiang, Y. Chen, Y. Yang, C. Wang, and K. J. R. Liu, "Dynamic Chinese restaurant game: theory and application to cognitive radio networks," *IEEE Trans. Wirel. Commun.*, vol. 13, no. 4, pp. 1960-1973, 2014.