

Bridging Separate Communities with Common Interest in Distributed Social Networks through the Use of Social Objects

D. Garompolo¹, A. Molinaro¹, A. Iera²

¹ Dept. DIIES, University “Mediterranea” of Reggio Calabria

² Dept. DIMES, University of Calabria, Arcavacata di Rende (CS), Italy
email: [david.garompolo, antmolin]@unirc.it, antonio.iera@unical.it

Abstract—In light of the growing number of user privacy violations in centralized social networks, the need to define effective platforms for decentralized online social networks (DOSNs) is deeply felt. Interesting solutions have been proposed in the past, which own the necessary mechanisms to allow users keeping control over their personal information and setting the rules to regulate the access of other users. Unfortunately, the effectiveness of this type of solutions is severely reduced by the fact that different user communities with a shared interest could be disconnected/separated from each other. This translates into a reduced ability in effectively spreading data of common interest towards all interested users, as it currently happens in centralized social networks. In order to overcome the cited limitation, this paper proposes a disruptive approach, which exploits the availability of a new class of Internet of Things (IoT) devices with autonomous social behaviors and cognitive abilities. Such devices can be leveraged as friendship intermediaries between devices’ owners who are connected to a DOSN platform and share the same interest. We will demonstrate that clear advantages can be achieved in terms of increased percentage of Interested Reachable Nodes (a specific measure of Delivery Ratio) in distributed social networks among humans, when enhanced with so called *Mediator Objects* adhering to the well-known social IoT (SIoT) paradigm.

Index Terms—Social Internet of Things, Distributed Social Networks, social mediator objects, content diffusion

I. INTRODUCTION

Recent scandals such as the one involving Cambridge Analytica have clearly shown that technology giants (Facebook, Google & Co, etc.) have serious problems in protecting user data. In order to guarantee users privacy protection and a greater control over their data, researchers have recently proposed several distributed platforms to implement decentralized online social networks (DOSNs) [1] either based on peer-to-peer (P2P) architectures [2], [3], [4], [5] or with a Web-based nature [6], [7].

In this context, an interesting decentralized platform for social Web applications, named Solid, was proposed by Tim Berners Lee in [8]. It decouples user data from applications that create and consume this data, ensuring they have a simple, generic and well-defined way to access data stored in the user’s Web-accessible personal online datastore (POD) [8]. This platform, in essence, allows users to maintain control over their personal information and to decide where it is stored, who can view it, and which applications can access it.

Web-oriented DOSNs can be, for example, developed by leveraging the Solid platform (Solid Social is an example) through which a user (*Source*) can diffuse her own content in a targeted manner to all people interested in receiving it. In Solid, the social graph visible to each user consists of the contacts stored in her POD, the contacts of these contacts, and so on. Contacts can be seen as an interface to manage the user’s distributed social graph. Interestingly, a user can either mark her contacts as public or make them accessible to a specific individual or group of people [8], [9]. While with a centralized social network the application knows the complete social graph, in the distributed case it only knows a partial graph consisting of (i) the contacts stored in the POD of the Source willing to disclose the content and (ii) the contacts in the PODs, which the Source is authorized to access. Obviously, authorization mechanisms in the use of the node’s contacts can be leveraged to guarantee a given content to be diffused to the interested nodes only. This allows reducing as much as possible the number of unnecessary visited nodes and making the information spreading as efficient as possible [10], [11], [12], [13].

Unfortunately, such an approach, although being very effective with a view to ensuring trust and security in accessing the information, at the same time amplifies the inherent limitations of decentralized social network solutions: *different communities with the same interest could very likely be disconnected/separated from each other, and some nodes potentially interested in a content could be isolated*. In other words, nodes that share the same generic interest cannot always be mutually discovered and cannot exchange contents related to that interest. This reduces the extent to which a Source is able to spread its content to the highest number of possibly interested nodes.

The associated risk is a decrease in the appeal of this type of solutions for the users playing the role of “prosumers” (i.e., both producers and consumers) of contents. As consumers, they could see the access to the content of their interest threatened by a discovery process hindered by fragmentation and possible isolation of some communities of interest. Even worse is the situation for a user who produces contents. In such a case, the producer would have difficulty in spreading her content to a large number of interested users and, if a

business is based on the delivery of such a content, the problem implies a reduction in revenues from commercial activities that operate on distributed platforms.

A new class of devices, showing autonomous social behaviors and having cognitive abilities, could help in addressing some of the highlighted issues. The idea is to allow such devices to act as “facilitators” of contacts and friendship among users with the same interests and, therefore, behaving as “bridges” between communities with similar interests which would not be connected in a distributed social network.

The main contributions that our research intends to provide are summarized below:

- the well-known Social Internet of Things (SIoT) paradigm [15][16] and the possibility that it offers to have a new class of devices, showing autonomous social behaviors and having cognitive abilities, is leveraged to increase the number of interested nodes that can be reached by a given content in a DOSNs based on the Solid platform;
- the new concept of *Mediator Object* is introduced, which refers to a device with a social and cognitive nature that mediates the content propagation towards other interested devices/users, otherwise unreachable. This concept could be seen as the first implementation, through Information and Communication Technologies (ICT), of the well-known concept of “social object” in sociological studies on object centered sociality [14], referring to an object creating a human connection between two individuals;
- the SIoT paradigm is evolved into an empowered version, named here *Enhanced SIoT*. It still exploits the basic management and control mechanisms proposed in the literature for the SIoT (included those related to Internet of Things (IoT) devices’ trustworthiness [17], [18], [19]), and adds new features to enhance the information diffusion among social devices, and thus to increase the number of interested nodes that can be reached by a given content (interested reachable nodes - IRNs);
- a set of analysis are conducted that show the advantages offered in terms of increased percentage of IRNs in a decentralized social network, when leveraging the introduced paradigms of Object Mediators and Enhanced SIoT, compared to the traditional case in which only friendships among humans are considered;
- a last additional side contribution is also the proposal of a possible methodology for modeling a SIoT network between devices starting from real human tracks in Social Networks, which can be useful to researchers in the field of SIoT, because today there are no tracks nor datasets from real SIoT devices available yet.

This paper is organized as follows. In Section II background details on the Solid system and the SIoT paradigm are provided. Section III summarizes the reference literature while Section IV addresses the proposed Enhanced SIoT, illustrates the concept of community and the role of the Mediator Object, and describes how discovery and diffusion take place in the

Enhanced SIoT. An analysis of the performance achieved by the proposed solution compared to a traditional centralized social network solution is the subject of Section V, while conclusive remarks are given in Section VI.

II. BACKGROUND

A. Distributed vs. Centralized Online Social Networks

A Decentralized Online Social Network (DOSN) is an online social network implemented on a distributed platform. By distributed we mean that all computing, storage and communication resources are provided by the users.

DOSNs are considered in many works in the literature as a possible solution to give users greater control over their data and, at the same time, to overcome typical problems of centralized online social networks, such as privacy, performance bottleneck, single point of failure, need to be online in all transactions, unexploited locality [1], [24], [25], [27].

While in the Centralized case the single service provider has control over all user data and could change the existing terms of service, in the Decentralized case there is a set of nodes that cooperate to guarantee all the functionalities. This gives to the users more control over their privacy [26].

Indeed, in a DOSN the user data will be stored locally in devices and controlled by end users instead of the OSN central entity (OSN provider). Consequently, the DOSNs can mitigate the privacy control issues, the problem of security and scalability and increase the flexibility and the ability to deal with big data problems [27] [28].

Another important problem of Centralized OSNs is the performance bottleneck due to the very high number of user requests and the huge amount of social data (all data exchanged in Social Networks regarding both user information and generated content). The DOSNs can alleviate the problem of performance bottleneck and avoid the single point of failure and the single point of attack.

Finally, centralized OSNs can suffer from other problems such as limited scalability and high maintenance costs to manage the data of so many users. In DOSNs, shifting to the user both the implementation of the infrastructure and the privacy and security control effectively reduces the operational cost [29],[30].

Table I synthesizes the main differences between centralized and decentralized OSN

B. Solid

Proposed by Tim Berners-Lee, Solid is a web-based open source platform that allows users keeping control over their personal information. Thanks to it, large social network companies will be allowed to use only part of the user data, and this permission can also be withdrawn at any time.

The Solid’s core is represented by the Solid POD, which can be seen as “a private website with data inter-operable with all apps”. It stores all the user’s personal information that will be linked from the outside in order to be used. In this model, only users will control their own information [8], [9]. The “Contacts” application manages a list of contacts stored on a

TABLE I
CENTRALIZED VS DISTRIBUTED OSN

	Distributed OSNs	Centralized OSNs
Computing, storage and communication resources	Provided by the users	Provided by the OSN central entity (OSN provider)
User data storage	Local data storage in devices controlled by end users	Remote data storage controlled by the OSN central entity (OSN provider)
Privacy and Terms of Service	Users have more control over their data privacy	The service provider has control over all user data and can change the terms of service
Security, Scalability, and Maintenance costs to manage users data	Mitigate security and scalability issues and increase the flexibility in dealing with big data problems	May suffer from security control and scalability issues
	Achieve an effective reduction of operational costs by transferring infrastructure, privacy and security control to users	May suffer from high maintenance costs to manage data from a huge number of users
Performance bottleneck	Counter the performance bottlenecks caused by high number of users requesting social data regarding both user information and generated content	Exposed to severe performance bottlenecks having to centrally manage the huge amount of social data and user requests
Single Point of Failure & Single Point of Attack	Avoid single point of failure and single point of attack, related issues	May suffer from single point of failure and single point of attack issues

user's POD. In Solid the user's social graph consists of the contacts stored in the POD, plus the accessible contacts of these contacts and so on; each user is identified by a WebID. The Contacts application maintains a set of vCards for the user's contacts by using the vCard ontology. Each vCard is a resource with a single Uniform Resource Identifier (URI) and can contain the user's WebID that it represents, in addition to other fields such as name and email. A user can mark a vCard as "public" or allow access to the vCard only to a single individual or to a specific group of people (identified by their WebIDs).

This new vision, which allows decoupling data from applications, brings with it new issues to be addressed in view of deploying social network applications for Solid (e.g., Solid Social). While in the centralized social networks the application (e.g. Facebook, Instagram, etc.), possessing all the data (in the Data Silos), knows the complete graph of the contacts of all users, in this new distributed vision this is no longer possible. In particular, each user will know, in addition to her contacts contained in her Solid POD, only the contacts contained in the Solid PODs of the users who have granted her permission to access their own PODs. Consequently, there will be a graph with many separate components (Local Knowledge).

Let us imagine that a user wants to disclose a given content in a targeted manner to all the nodes that are interested in receiving it. In this case, being the social network application distributed and therefore no longer aware of the complete graph, during the discovery phase it might not be able to reach all the nodes interested in receiving the content, as it was in the centralized case. Consequently, even during the targeted diffusion phase (i.e., diffusion to the interested nodes only) a node may not be able to spread its content to all the other interested nodes.

C. Social IoT (SIoT)

The IoT is a paradigm that in recent years has received much attention from the scientific community. A possible way to look at the IoT is to define it as *"a conceptual framework that leverages the availability of heterogeneous devices and inter-connect solutions, as well as augmented physical objects that provide a shared global information base to support the design of applications that involve both people and representations of objects"* [21]. Accordingly, every person and thing in the IoT has a virtual counterpart that can be localized, addressed, and readable in the Internet (in the Cloud, the Fog, or at the network Edge).

Objects act as prosumers of services and collaborate with other counterparts to reach common goals. This leads to design a new generation of "smart objects" that will have to operate in complex contexts; it is unlikely that the single object will have the capacity to deal with such a complexity. Like some species of animals that, to cope with complexity and the difficulties of the environment in which they live, have created a dense network of social relationships, in the same way a new generation of social objects has been envisaged to form an augmented IoT, called SIoT, that applies to IoT the typical concepts, solutions, and technologies of social networks [22]. Accordingly, an IoT object becomes part of and acts in a social community of objects and devices. In this context, the social networks of objects are built among objects that are owned by human beings, who may have no connection among them [16].

The application of the social networking principles to the IoT (i.e., SIoT) can bring several advantages in network navigability, scalability, and discovery of objects and services. Moreover, powerful models of trustworthiness management

designed for social networks can be reused to address IoT-related issues.

In SIoT a basic set of inter-device relationships are defined: parental object relationship (POR), established between objects belonging to the same production batch; co-location object relationship (C-LOR), established between objects used in the same location; co-work object relationship (C-WOR), established between objects that often cooperate together to provide a common IoT application; ownership object relationship (OOR), established between objects belonging to the same user, and social object relationship (SOR), established between objects that often come into contact because their owners come into contact with each other during their lives.

The creation and management of such relationships can take place without human intervention [23], but always strictly complying to rules set by the devices' owners. Several further types of friendship have been added to these relationships over the time, driven by specific application environments (e.g., the social Internet of vehicles).

III. RELATED WORKS

A. DOSN classification

In [28] DOSNs are classified into two categories: Web-based and P2P-based DOSNs. The first category, which is considered as a basis in our work, is characterized by a distributed web server infrastructure, and includes systems such as Diaspora [7] and "Friend-of-a-Friend" (FoaF) [6]. These solutions need web space or deploying web servers. Users can publish their profiles in their Web space and manage access control rules (access authorizations) locally, to specifically allow the recovery of attributes and resources reserved for the selected users. Web links to other users' profiles are used to represent the Contact List and thus recreate the social graph.

The second category includes systems such as Likir [2], Peerson [3], Safebook [4], which exploit the advantages of the P2P principle to allow the publication, search, and retrieval of profiles and their attributes, very similarly to conventional P2P file-sharing systems. In them, the resources are kept locally and the profiles are stored in the local devices instead of the Web, and controlled by the users themselves. In P2P-based DOSNs one of the main challenges concerns the replication of user profiles, which must always be available online even when the user is offline. The issue of data availability/persistence is addressed in [1].

In summary, while web-based systems rely on a dedicated web space where user profiles can be stored and retrieved, P2P-based systems exploit only the local and shared resources of the P2P overlay. Obviously, exploiting the rather unreliable storage services of other peers, which are subject to churn (i.e., nodes entering and leaving the network, or changing state from online to offline continuously), requires more sophisticated means to keep the data available, which in turn causes overhead and higher implicit costs shared among the participants. Differently, in our research we follow a Solid-like web-based system design and, since the PODs in which the

profiles are stored are always available online, there is no need for replicas of the same profile.

B. Content sharing and diffusion

Our reference scenario is characterized by the presence of Interest Communities (i.e., sets of nodes that share the same interest) that are separated, disconnected, and unreachable from each other. This represents the problem we aim to solve.

The reasons for which the concerned nodes can be unreachable are the most disparate and vary according to the considered scenario. They can be related (i) to the overlay topology; for example, in Friend-to-friend (F2F) networks, communications can only take place between "friends"; (ii) to the type of content diffusion; for example, in the case of targeted diffusion, the content is sent only to the interested nodes without any bridging performed by non-interested intermediate nodes; or (iii) to the fact that the interested nodes are located in different network Partitions (i.e., areas of a network that are connected in the presence of a mobile node and disconnected in its absence) caused by churn phenomena, for example. This is not an exclusive feature of DOSNs, but it may also occur in other systems, ranging from mobile ad-hoc to opportunistic networks.

Several methods have been proposed in the literature that consider the community structure to be a significant property of social networks and propose methods for community identification, most of which apply to static networks. There are also interesting studies available, as the one described in [35], focusing on dynamic and overlapping community detection. There, the authors propose a novel method able to track the evolution of overlapping communities in dynamic social networks based on topology potential field, which jointly solves the problems of overlapping community detection, dynamic community identification and community evolution analysis. Through this method dynamic overlapping social networks can be accurately partitioned and all kinds of community evolution events efficiently tracked.

In general, the goal of a content-sharing system is to move content items to devices owned by users who want to access such a content [34]. In our reference context the goal is very similar, as a Source node wants to spread its content to the highest possible percentage of interested reachable nodes. The main issue in such a scenario is that, when diffusing the information relevant to the Source's interests both to the interested and not interested nodes during the Discovery phase, not all nodes will be equally cooperative and equally reachable. Also, they will not be equally willing to authorize the access to their PODs and to re-forward information on behalf of the Source to their Contacts, otherwise unreachable by the source directly. The solutions proposed in the literature to spread contents to interested users belonging to different Communities of Interest/Partitions are manifold and usually depend on the type of scenario and related assumptions. In [34] an ad-hoc content-sharing system for mobile devices is proposed within an opportunistic network scenario. In it, mobile devices share contents and interests and can store-

carry-forward contents on behalf of other nodes, based on interests, therefore connecting otherwise disconnected devices. In [34], Huggle introduces a content delegation mechanism that allows to selflessly disseminate a given number of items based on the interests of other nodes (third-party nodes). This is particularly important in the presence of network Partitions. The mobile nodes that provide this type of connectivity are called “data mules”, reflecting the idea that they carry data between otherwise partitioned areas of the network, [36]. In [34] it is clearly explained that depending on the network structure and the users’ interests, a content may not reach the interested nodes without exploiting data mules that carry the content although they are not interested in it. Although the scenario is very different from ours, it presents a problem very similar to the one we intend to address. Here too, the goal is to spread the contents to devices owned by users who need such contents [34].

Several studies aimed to design valid mechanisms to diffuse the content as much as possible to the interested nodes. For example, [12] analyses how maximizing the total weight of the content receivers, which is measured in proportion both to how much the users themselves are interested in the content and based on their ability to connect with other interested users.

There are many reasons why nodes should decide to spread the information only to the interested nodes. First, the latter are more active during the diffusion process [13]. The second important reason is the guarantee of greater privacy and security. So a possible research goal is to minimize the number of nodes not interested that are involved in the diffusion process [10], [11]. In our case, even in the Discovery phase, when information on the Source interests is diffused, we are able to guarantee privacy by sending it totally anonymously.

In [37] the goal is to make the information available in those regions where there are interested users without overusing the resources, i.e., avoiding flooding. Third, it must be considered that the interested nodes are also the desired receivers, and therefore they are motivated to participate in the diffusion process to receive the content of interest. Finally, users with similar interests also have a higher frequency and probability of communicating with each other, which allows a more efficient diffusion [13]. This feature is exploited to improve content diffusion in [38]. In particular, a user will download the content from another user she meets if and only if (i) the topic of the content is of her own interest, (ii) it interests her friends, or (iii) it interests the users she meets.

In our framework the content of the Source is diffused *only to interested nodes* and, consequently, there will be no uninterested data mules that connect the separate Communities of Interest (Partitions). Differently, during the Discovery phase the information on the Source’s interests will be sent (anonymized) *both to interested and not interested nodes*. Intermediary nodes (even if not interested) spread the information relevant to the Source’s interest and allow the Source to reach nodes that otherwise it would not be able to directly reach. As a result, on the one hand, during Discovery, the Source is able to indirectly reach a higher number of interested

nodes, thanks to the presence of intermediaries. On the other hand, during Diffusion, the Source manages to send its content as efficiently as possible to the interested nodes only.

Privacy is guaranteed both during the Discovery phase, thanks to the anonymity of the information disseminated, and, during the Diffusion phase, by sending the content directly to the desired receivers only.

Due to the churn in the Social Overlay (SO), a limited set of links may be available for reconfiguration and cause transient network partitions, which are responsible of long unacceptable delays in the content diffusion phase [25], [39]. As a solution to this issue, a hybrid architecture is proposed in [25] that on the one hand exploits the SO for fast, decentralized and friend-to-friend communications, but occasionally exploits the access to the cloud to overcome the high delays caused by the purely decentralized solution. Comparing it with our work, we see the analogy between the transient network partitions and the separate Communities of Interest, but the focus is slightly different because, while in our case a Source node wants to spread a content to all the interested network nodes, the goal in [25] is to allow efficient profile-based communication between direct friends. In both cases, however, the problem lies in the limited set of available links. Unlike the solution presented in [25], in which the concept of purely distributed architecture is lost, in our case we want to show the advantages in terms of reachability that are obtained by extending the set of friendships through social object relationships (SIoT Contact List), while still maintaining a distributed approach to reach any interested node belonging to other communities.

IV. ENHANCED SOCIAL INTERNET OF THINGS

A. Community

Fig. 1 shows two Interest Communities in our reference scenario. Let us assume, for example, that Community *A* is made up of nodes that share at least one common interest, e.g. the “Football” Interest. This does not mean that nodes belonging to a community cannot have other interests in common with nodes belonging to other communities, but, to simplify, in Fig. 1 we do not illustrate this case. In the remainder of the section, let us assume that:

- C_A = Community A = Community of nodes that share at least the “Football” Interest = $C(I_A)$;
- I_A = Interest A = Interest “Football”;
- $V(I_A)$ = set of all nodes (vertices of the graph) with I_A ;
- $C_A \subseteq V(I_A)$.

Each direct dashed line between a pair of nodes indicates that both nodes can be mutually reached by performing a Search for that specific interest. They can be reached using the “Contacts of Contacts” mechanism, as they are authorized to access the contacts of their contacts. If two nodes within the same community are not connected by a dashed line, it means that they cannot be reached via direct search, i.e., if one of the two nodes searches for that interest, the other’s vCard will not be returned.

Taking up the concept of Reachability from Graph Theory we know that: a node n_2 is reachable by a node n_1 (or n_1 reaches n_2) if there exists a path from n_1 to n_2 . We define two relations on the set of nodes:

- (a) the reachability relation $R \rightarrow$ such that $n_1 R \rightarrow n_2$ if n_2 is reachable from n_1 ;
- (b) the mutual reachability relation $R \rightleftarrows$ such that $n_1 R \rightleftarrows n_2$ if $n_1 R \rightarrow n_2$ and $n_2 R \rightarrow n_1$.

At this point, considering that in our scenario the mutual reachability relationship indicates that “*both nodes can be mutually reached by performing a Search for that specific interest and they can be reached via the “Contacts of Contacts” mechanism, as they are authorized to access the contacts of their contacts*”, we can represent this relationship with the following formalism:

$n_1 R \rightleftarrows n_2$, which graphically, in Fig. 1, corresponds to the direct dashed line between n_1 and n_2 , with $n_1, n_2 \in C_A$

A node of a community that intends to spread the content belonging to a specific interest is called Source. This node, thanks to the permissions received from its contacts to view their contacts, can directly reach with its content the interested nodes. The reachable nodes are defined as Direct Interested Reachable Contacts (D-IRC) of the Source. Furthermore, its contacts may relaunch the search to reach further interested nodes (otherwise not reachable from the Source) by operating as Sources. This hopefully allows the Source’s content to reach all the nodes interested in that specific interest (the respective IRCs), which for the Source are to be considered Indirect IRCs (I-IRC).

We must now distinguish two types of relationship:

- $DR \rightarrow$ = Direct Reachability

It holds if a node, thanks to the permissions received from its contacts to view their contacts, can directly reach with its content the interested nodes; in this case, the set D-IRC of the Source node S is expressed as:

$$D-IRC(S) = \{\forall n \in V(G) \mid S DR \rightarrow n, n \subseteq V(I_A)\}$$

given that:

$d_A = d(I_A)$ = data (content) characterized by Interest A;
 S = the Source that wants to disseminate d_A ;
 $S DR \rightarrow n \in D-IRC(S)$;
 $V(G)$ = set of all nodes of the graph G ;

and represents the set of nodes, directly reachable by S , interested in the content characterized by Interest A

- $IR \rightarrow$ = Indirect Reachability

It holds if the node’s contacts may relaunch the search to reach further interested nodes (otherwise not reachable from the Source) by operating as Sources. In this case

The set I-IRC of the Source node S is expressed as:

$$I-IRC(S) = \{\forall n \in V(G) \mid S IR \rightarrow n, n \subseteq V(I_A)\}$$

given that:

S = the Source that wants to disseminate d_A ;
 $S IR \rightarrow n \in I-IRC(S)$;

and represents the set of nodes, indirectly reachable by S , interested in the content characterized by Interest A

By applying this process recursively, a Source node will be able to disclose its content to all interested nodes of a community (both D-IRCs and I-IRCs); specifically, all the reachable nodes will constitute the community.

Let S be the Source that wants to disseminate d_A . The set of nodes directly reachable from S plus that of nodes indirectly reachable from S , interested in the content characterized by Interest A, is equal to the set of all the nodes reachable from S , interested in the content characterized by Interest A.

We call this set $C_A = D-IRC(S) + I-IRC(S) = IRC(S)$,

defined as:

$$C_A = \{\forall n \in V(G) \mid S DR \rightarrow n \parallel S IR \rightarrow n, n \subseteq V(I_A)\}$$

Finally, we assume that if two nodes sharing the same interest are in two different communities, then none of the nodes in the first community can reach any node in the second community. In Fig. 1 the node marked as “interested node” is a node that cannot be reached by searching from any node belonging to Community A, otherwise it would belong to it, and therefore it is part of Community B.

In order to enable content diffusion also to interested nodes belonging to other separate communities, we propose here to exploit the SIoT concept, as explained in the following subsection.

B. Mediator Object

The idea of a Mediator Object is leveraged in this context to reach otherwise unreachable interested nodes; its objective is precisely to “mediate” the content propagation from Community A, wherein the content is generated by the Source node, to a separate Community B, where the interested node is located. This sort of “bridging” between two communities can be achieved through objects that have a social and cognitive behavior and interact with one another.

In Fig. 1, we consider a SIoT network between the two communities made up of cognitive objects. In the remainder of the paper, by “*cognitive object*” we will indicate an object with the ability to proactively search the social network of objects to which it belongs (through the use of the SIoT platform) and to understand, from the Interest Descriptors (defined in the next Section) and from previous events, whether “friend objects” from other communities may be interested in receiving a certain content. The way it happens will be described in the following. Obviously, in order to carry out

its functions of mediation between communities, the Mediator Object must be a cognitive object.

An exemplary use case is given by a cognitive object which understands, through the mechanism described below, that the news circulating in Community *A* of soccer fans may also interest another “friend” and “trusted” object, whose owner belongs to Community *B* of soccer bettors, for example. We assume that a social object relationship was indeed created between the devices, according to the SIoT rules. The devices became friends, because, for example, they often came into contact in a soccer stadium, although their owners do not know each other.

In this context, a first problem is linked to the way in which a cognitive object can know if “friend objects” from other communities may be interested in receiving a certain content. To this end, each device is associated with an Interest Descriptor, i.e. a vector of words (keywords) that describe the interests of its owner.

There are several ways to derive the Interest Descriptor, depending on the information available. Without losing generality, we can refer to an exemplary solution based on the Visual User Interest Profile (VUIP), as in [40].

In order to better understand what the VUIP is, it is necessary to introduce the concept of user profiling, which can be defined as the process of identifying data relating to the user’s domain of interest. A device can infer the owner’s profile based on a set of images, accessed through that device, that describe her interests (as in Instagram, Facebook, etc.). For example, by leveraging deep learning techniques, from the images it is possible to obtain the corresponding VUIP (that is a vector of keywords derived from images) that can be used as an Interest Descriptor.

The method we will adopt is just one of several possibilities allowing to extract the keywords (tags) associated with images. As an example, another work that deals with Image Tagging problem, i.e. the extraction of tags from the image, is described in [33]. The authors describe a new Deep Collaborative Embedding (DCE) model for social image understanding applied to social image tag refinement and assignment, content-based image retrieval, tag-based image retrieval and tag expansion.

During the discovery process, better described in the remainder of the paper, the device itself can send its VUIP (coinciding with the VUIP of the owner) to its first social neighbors in the SIoT, i.e., nodes with which it has already established at least one social object relationship. We assume that each device has such a capability of profiling the interests of its owner and creating the corresponding VUIP.

Obviously, during the whole process of browsing the SIoT, the Source’s VUIP exchanged among the devices *must remain anonymous*.

To our purposes, we also define a new social relationship between devices, the *Co-Interest Object Relationship (C-IOR)*. Such a relationship is established between two devices when the VUIPs of their device owners are sufficiently similar, i.e., when the degree of similarity between the VUIPs exceeds a certain threshold.

C. Basic Mechanism for the C-IOR establishment

In general, an Interest Neighbor of a *S* device is a device connected via C-IOR to *S*, therefore it is a device sharing one or more interests with *S*. We describe in the following the basic mechanism that characterizes the phases of Interest Neighbor discovery and corresponding C-IOR establishment. For greater clarity we will refer to Fig. 1.

The *S* device derives its owner’s VUIP. *S* sends its VUIP to all its first social neighbors. The first social neighbors of *S* are the nodes with which *S* has already established at least one SIoT relationship (such as C-LOR, SOR, etc.); among them there can be also cognitive objects, including the Mediator Object *M*. We assume that all the devices are cognitive. The cognitive objects can disclose the *S*’s VUIP also to cognitive objects belonging to other communities. Every first social neighbor of *S* who receives the VUIP of *S*, including *M*, checks whether it is possible for it to establish a C-IOR with *S* (based on the degree of similarity between VUIPs) and sends in turn the *S*’s VUIP (appropriately anonymized) to its own first social neighbors who have not received the *S*’s VUIP yet.

The VUIP of *S* is recursively delivered with a maximum Time to Live (TTL) of 6 hops (small world network property) set by the Source and decremented hop-by-hop.

If an interested node in another community, e.g., the *T* node in Community *B*, based on the similarity between its own VUIP and the received *S*’s VUIP, decides to establish a C-IOR with the owner of the received VUIP (it does not know that *S* is the owner), then the C-IOR establishment request of *T* (which includes *T*’s identity) goes backwards, forwarded by the intermediate nodes, until it gets to Source *S*. The reason why *T* cannot directly send a C-IOR establishment request to *S* but rather the establishment request of *T* is brought back to the Source by leveraging the intermediate nodes is that the VUIP owner (i.e. *S*) must always remain anonymous in order not to infringe *S*’s privacy.

In short, to ensure the privacy of the Source, each intermediary knows only the identity of the “previous node” that forwarded the VUIP to it, and to which it will have to forward any request of C-IOR establishment coming from the interested node *T*. After 6 hops, the VUIP expires and it is no longer forwarded. Once the C-IOR establishment request of the concerned node reaches the Source, a C-IOR can be established that directly binds node *T* and node *S*.

The Cognitive devices, the Mediators in particular, by mediating the propagation of the VUIP (Interest Descriptor) of *S* from one community to another, and by enabling the establishment of the C-IOR between nodes belonging to different communities, allow the propagation of content across separate communities.

The described mechanism, based on the VUIP propagation, is obviously general and can be implemented with any type of Interest Descriptor obtainable for a given user.

D. Enhanced Discovery and Enhanced Diffusion

The foundation of the Solid’s “Contacts” is the “Contacts of Contacts” mechanism, in line with the more known “Friends

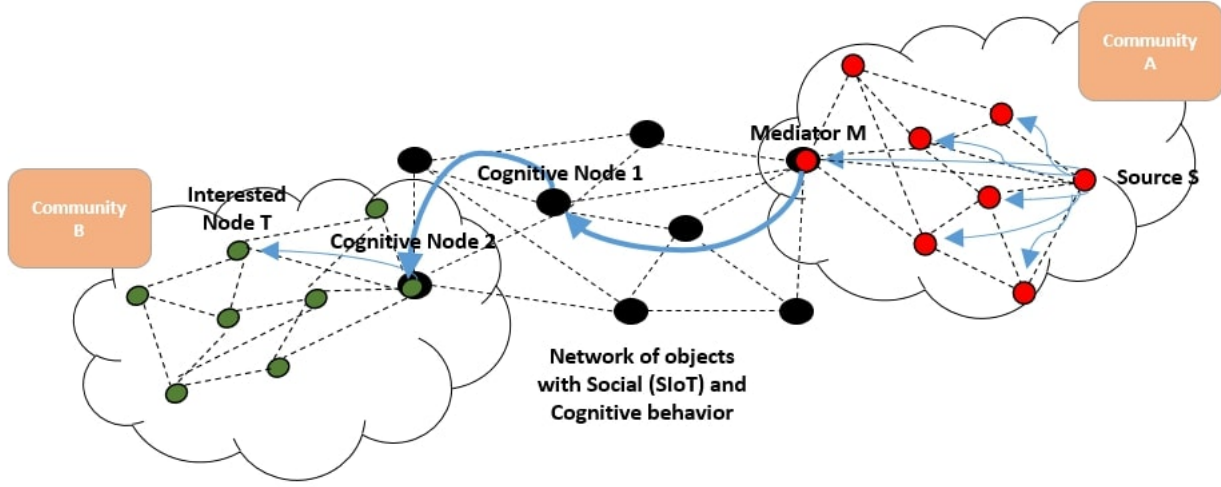


Fig. 1. Basic Mechanism for the establishment of the C-IOR.

of Friends” mechanism. Accordingly, each node performing Discovery (search) can scan, in addition to the Contacts (friends) contained in its User POD, also the Contacts of its own Contacts, which have granted it authorization to access its Users’ PODs. This is possible thanks to the link-following SPARQL [8], [9].

We are assuming that in the User POD each user keeps her own list of contacts (*Contact List*). In addition a user can scan the Contacts, contained in the User PODs of other users (her contacts), to which it is authorized to access.

On the other hand, the list of nodes, which each device is linked through SIoT relationships (*SIoT Contact List*) is locally stored on the device itself, and specifies the type(s) of relationship(s) through which the nodes are mutually linked. More details on how to create a distributed SIoT network are illustrated in the literature [41].

During the Discovery phase, a device will scan: (i) its User POD (i.e., its own Contact List), (ii) the User PODs of its Contacts that authorize it to access, (iii) and the User PODs of the Contacts reachable recursively (through the chain of authorizations) that authorize it to access. In addition, the device will know its own SIoT Contact List. Specifically, the node will scan both the interested nodes and the nodes not interested in the content. The result of the Discovery will return only the interested nodes among the scanned ones.

For brevity we call them ED-IRC (Enhanced Direct Interested Reachable Contacts). We use the term “Enhanced” when we leverage also the SIoT Contacts.

In short, with *Direct* we identify nodes reachable in the Discovery phase, therefore directly reachable by the node that wants to disclose the content, while with *Indirect* we mean the nodes that cannot be reached in the Discovery phase, but are anyway reachable in the Diffusion phase through intermediate nodes, which do not authorize (in the Discovery phase) the Source node to access their User POD.

After each node has performed Discovery and knows its ED-IRC, only Interested Nodes will be considered during the Diffusion phase. A node that wants to disclose content can disclose it directly to its ED-IRC.

V. PERFORMANCE ASSESSMENT

A. Generation of realistic SIoT datasets

We followed the procedure used in [42] to obtain a reliable dataset on the meetings (co-locations) that take place between people. Accordingly, by starting from the Check-ins Dataset (Brightkite [43]) we obtain a co-location if and only if two Check-ins of two different users took place within 250 meters and within 1800 seconds.

While the mentioned procedure refers to contacts between persons, we are interested in the contacts between devices that may bring to the triggering of social relationships between them. Therefore, without losing generality, we assume that each person takes a mobile device with her and leaves a fixed device at home (her Home-Point). The assignment of a given model to a device (useful for establishing POR relationships) is carried out based on the ownership report of the Global Web Index in 2017 [45] calculated on 50000 users.

In this way, by replacing each user with their own devices, we can easily obtain the meetings that took place between devices, from those that took place between their owners. Like in the aforementioned paper [42], we will only consider users with at least 10 Check-ins and with at least 10 different Check-ins places. This allows us to exclude less active and therefore less interesting users.

Various methods are proposed in the literature to derive the positions of the *Home-Points* of the users [43], [46], [47]. We recall here that the Home-Points are fundamental to derive the position of the fixed devices that are part of the SIoT network. The method we used to calculate the users’ Home-Points is the one proposed in [47].

In general, there are places that we call for convenience *Points (or Places) of Interest* (PoI), towards which nodes will be more likely to go. We can see them as places of particular importance, places where specific activities are carried out, where people cultivate certain interests. Around these places we will have a greater concentration of meetings (in the reminder also referred to as co-locations).

Each meeting will take place more or less near a PoI. Thus, we assign a specific PoI to each meeting. This means that if a meeting took place near a certain PoI, with a certain probability the user (and her relevant device) has the Interest associated to that specific place. In particular, to assign an Interest to a co-location, we used a Foursquare dataset [48], [49], [50] that associates each PoI (in terms of latitude and longitude) with an Interest. In practice, by putting into relationship the meeting position and the PoI position in the Foursquare dataset, it is possible to assign a relevant Interest to each co-location.

As for the Interests considered in our experiments, we started with the Foursquare Interests and we had to group them into Macro-categories, because they were defined in a very specific way. This had the effect that a large enough number of Communities with common Interest were not created to guarantee us a good statistical confidence in the results of our analyzes.

Each Foursquare Interest is described by a single keyword, while a Macro-category (Interest) is made up of a set of keywords. The interest associated with a user (or device) and also with a content will be a Macro-category (Interest). All Foursquare's Interests fall into 52 Macro-categories that we call Interests. As an example, the Macro-categories we used in the performance evaluation studies illustrated in the remainder of the Section are: Sweet Food (Interest 3) including the Foursquare Interests: {'Pastelaria', 'Ice Cream', 'Yogurt', 'Donut', 'Dessert'}; Italian Food (Interest 4) including the Foursquare Interests: {'Meatball', 'Wine', 'Pizza', 'Ice Cream'}; Café Bar (Interest 6) including the Foursquare Interests: {'Bistro', 'Breakfast', 'Cafe', 'Tea Room', 'Donut', 'Dive Bar', 'Cupcake', 'Coffee', 'Bar'}.

Obviously, a meeting near a certain PoI could happen casually. To understand if the user assiduously attends that PoI, a given number of meetings must take place near that place, or better, *near places of that type*. Therefore, we set a threshold on the minimum number of meetings near a PoI (set to 10 in the shown performance campaign) required to assign that Interest to the user. In this way, from the co-locations (meetings) it is possible to obtain the PoI frequented by people, from which we can obtain their *interests*. Each device is associated with an Interest Descriptor, i.e. a vector of words (keywords) that describe the interests of its owner.

Following the described procedure we are able to establish the *Communities*. In particular, from the Brightkite Dataset [43] we get the *Friendships* between people and we establish the *Authorizations to Access Contacts* in Solid PODs. In this way, assuming that there is a Source that wants to spread content related to a given Interest, we are able to obtain the

Community.

At the same time, from the co-location events (meetings) we establish the *SIoT relationships* according to the SIoT rules available from the literature [15].

B. Use cases

In the studies presented in the remainder of the paper, the objective is to compare the mean IRN percentage (ie, the percentage of interested nodes reachable) obtained by using the *Enhanced SIoT* mechanism, which leverages the basic mechanism for the establishment of C-IOR, with the one obtained by the *Friendships* mechanism, in which only Brightkite friendships are leveraged.

More specifically, in the Friendships case, each node will be able to diffuse the Source content to all the interested nodes contained in its own Contacts List, in the Contact Lists of other users (her contacts) which it is authorized to access. In the Enhanced SIoT case, each node will be able to diffuse the Source content to the same interested node of the previous case, with the addition of those contained in its own SIoT Contact List.

A sufficient number of simulations were performed to be able to obtain statistical confidence in the IRN values shown in the curves.

C. Assumptions

The following assumptions hold:

- All SIoT relationships are considered except the C-WOR, which as demonstrated in the article [51] has a negligible contribution in terms of navigability [44].
- A threshold of 10 Check-ins is set in a specific type of PoI for the assignment of the relevant interests of a user.
- Each person brings a mobile device with her and leaves a fixed device at her home.
- In the Enhanced SIoT case (with C-IOR Basic Mechanism described in Sec. IV-C) node *A* spreads the Source data to node *B* of another community if and only if:
 - the two nodes are connected via a SIoT relationship or via a SIoT relationships path (connection in the social graph of devices);
 - the VUIPs of the two nodes have a similarity higher than a certain threshold (Cosine Similarity ≥ 0.5 [40]). The first two conditions imply the establishment of a C-IOR between the two nodes;
 - the *B* node has the specific interest of the data (which the *A* node wants to spread) in its own VUIP. The third condition implies the presence of a C-IOR between the two nodes associated to such specific interest.
- Each node that has SIoT relationships with nodes belonging to other communities (communities other than its own) acts as a potential Mediator.
- Scenarios as realistic as possible are considered. A limit is set on the number of hops for the diffusion of the Source's VUIP (TTL) in the Discovery phase, as it is more realistic to assume that not all nodes are willing to spread the

VUIP on behalf of another node. The percentage of nodes that spreads the Source VUIP to the different hops is varied during the simulations. In addition, since it is objectively less likely that a node makes its contacts available when increasing the social distance, then the percentage of nodes that provide authorization to access their PODs is assumed lower as the number of hops increases.

- It is assumed that every Source that spreads its own content will spread it to all possible interested nodes. In particular, in the Friendships case it will spread it to all the interested nodes contained in its own Contacts List and in the Contact Lists of other users (her contacts) which it is authorized to access. In the Enhanced SIoT case interested nodes in its own SIoT Contact List are also considered.
- Without losing generality, we assume that unless otherwise indicated we consider Interest 3 (“Sweet Food”) and the related Communities.
- It is assumed that not only interested nodes, but also not interested ones can authorize access to their PODs.
- Unless otherwise indicated, all nodes, including isolated nodes, are considered.

D. Performance by varying the number of nodes that spread the Source’s VUIP

The aim of the first performance evaluation is to investigate how the mean IRN percentage varies when varying the percentage of the nodes that diffuse the VUIP of the Source at each hop, by keeping fixed the percentage of nodes that authorize access to their PODs. The nodes that spread the Source VUIP are the nodes that act as intermediaries, allowing the Source to reach Contacts otherwise unreachable. The reported results consider a percentage of the nodes that spreads the Source’s VUIP at each hop equal to 100%, 90%, 60%, 30%, and 10%, and a number of hops for the VUIP diffusion equal to 4. All simulations were carried out in order to obtain a high statistical confidence (95%).

In Figure 2 the solid curves represent the trends obtained when exploiting all the social object relationships in the Enhanced SIoT case. The dotted curves represent the trends obtained if only Brightkite friendships are used (Friendships case). It is assumed that the percentage of nodes that authorize access to their POD is 100% at the first hop.

By observing Figure 2 we can appreciate the higher values in terms of mean IRN percentage obtained in the Enhanced SIoT case compared to the Friendships case. This means that through the Enhanced SIoT it is possible to reach a greater number of interested nodes. This is due to the presence of SIoT relationships and of all the additional proposed features and mechanism previously described, from the Mediator object to the basic establishment mechanism for the C-IOR.

The first two hops are those that have a greater increase in terms of mean IRN percentage (greater slope). We can note also the faster convergence in the Enhanced SIoT case compared to the Friendships case. This does not only mean

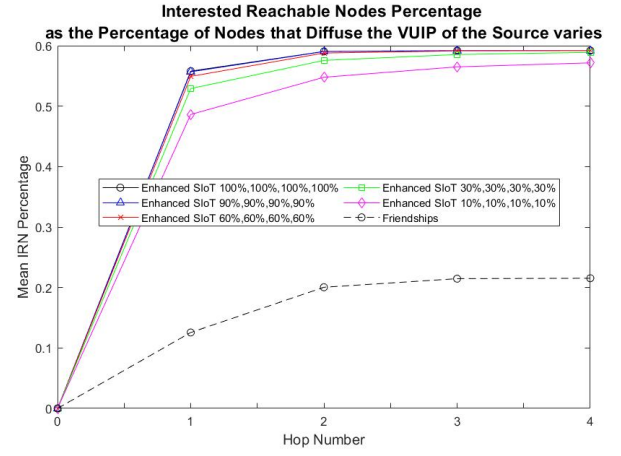


Fig. 2. Mean IRN percentage as the percentage of nodes that diffuses the Source VUIP at the different hops varies (Enhanced SIoT case vs. Friendships case).

that through the Enhanced SIoT a greater number of interested nodes can be reached, but also that they can be reached in a lower number of hops.

By observing Figure 2 it also clearly emerges, as we expected, that the obtained values in terms of mean IRN percentage increase with the increase in the percentage of nodes that diffuse the Source’s VUIP, and with the increase in the number of hops. We can note that also in the worst Enhanced SIoT case (in which only 10% of the nodes diffuse the Source’s VUIP at each hop), higher performance levels are achieved with respect to the Friendships case.

The low values obtained in general depend on the high number of interested isolated nodes present in the network for the specific scenario. As the number of hops increases, the increase in terms of mean IRN percentage becomes smaller, because most of the interested nodes that can be reached have already been reached.

E. Performance as the percentage of nodes that authorize access to their PODs changes

The second study aims to investigate how the mean IRN percentage varies with the percentage of nodes that authorize access to their PODs at different hops, by keeping the percentage of nodes that spread the Source’s VUIP fixed. Let us consider the limit of 4 hops, in which there will be nodes authorizing the access to their PODs. The labels of Figure 3 report the percentages of nodes that authorize the Source to access their PODs in each of the 4 hops. The reason why in Figure 3 we have set decreasing percentages of nodes that authorize the Source to access their PODs at each hop is that it is correct to assume that friends are more willing to authorize access to their PODs than friends of friends and so on. According to the real social dynamics in networks, the more socially distant one node is from another one, the less likely this node will authorize this latter node to access its POD.

Again, in Figure 3 the solid curves represent the trends obtained when all the social object relationships are considered in the Enhanced SIoT case. The dotted curves represent the trends obtained if only Brightkite friendships are used (Friendships case). We assume that the percentage of nodes diffusing the Source VUIP is 100% at the first hop, i.e. all the nodes spread the Source's VUIP.

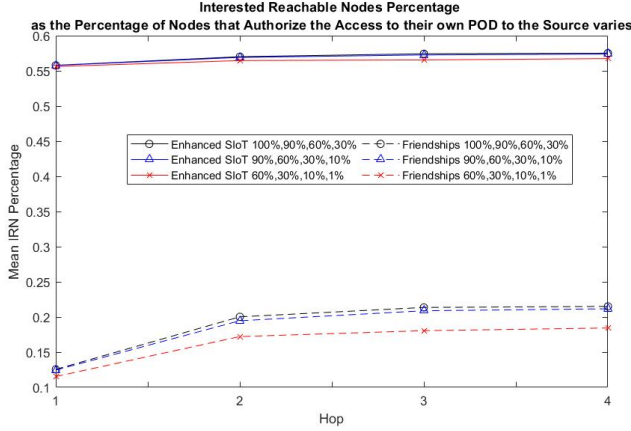


Fig. 3. Mean IRN percentage when varying the percentage of nodes authorizing access to their POD at the different hops (Enhanced SIoT case vs. Friendships case).

From Figure 3 we can note the higher values in terms of mean IRN percentage obtained in the Enhanced SIoT case compared to the Friendships case. Also here, the Enhanced SIoT is able to reach a greater number of interested nodes. By observing Figure 3 it also clearly emerges, as expected, that the obtained values in terms of mean IRN percentage increase with the increase in the percentage of nodes that authorize the access to their PODs, and with the increase in the number of hops (in which there are node that provide authorization to access their PODs to the Source). We can note that also in the worst Enhanced SIoT case, higher performance levels are obtained with respect to the Friendships case. Here, again the low values in general depend on the high number of interested isolated nodes present in the network. The reader can note that the gain obtained with a higher percentage of nodes, which authorize to access their contacts, is more accentuated in the Friendships case than in the Enhanced SIoT case. Also, the first two hops are those that show a greater increase in terms of mean IRN percentage (greater slope of the curves). This is due both to the fact that with the increase in the number of hops, most of the nodes that can be reached have already been reached, and to the fact that in the first hops we set higher percentages of nodes authorizing access to their PODs. This latter assumption has not to surprise, because it is correct to assume that friends are more willing to authorize access to their PODs than friends of friends and so on. The more socially distant a node is, the less likely this node will authorize access to its POD.

F. Performance by varying the kind of SIoT relationships between devices

A further objective of our study is to observe how the mean IRN percentage changes when the combination of SIoT relationships vary. For this purpose, simulations have been conducted in which six different combinations of SIoT relationships are considered. Figure 4 shows the variation of the mean IRN percentage, assuming that the 100%, 90%, 60%, and 30% of nodes respectively spreads the VUIP of the Source (act as intermediaries), in the Enhanced SIoT case.

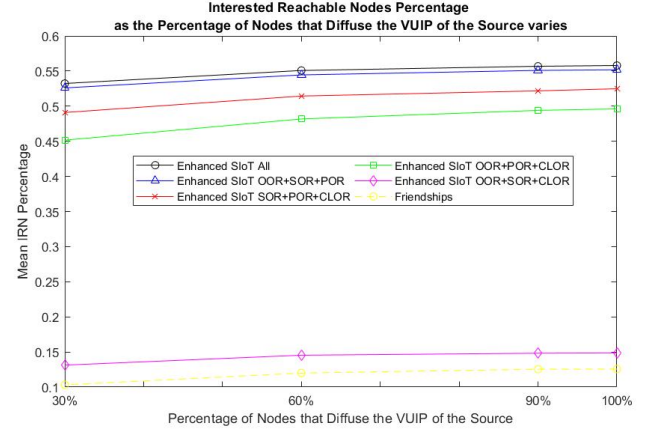


Fig. 4. Mean IRN percentage for different combination of SIoT relationships, as the percentage of nodes that diffuses the Source VUIP changes (Enhanced SIoT vs. Friendships).

A first evident result is that POR is clearly the social object relationship that weighs most on the obtainable mean IRN percentage values, followed by the SOR, the OOR, and the C-LOR. POR friendships in fact depend only on the model of the device and are often relationships that connect devices that are very distant from each other and belong to different communities. Given their characteristic of being "long-range" relationships, the relevant role, confirmed by the curves, in connecting users belonging to different communities otherwise separated was expected. The advantage in terms of the considered metric that the Enhanced SIoT case offer compared to the Friendships case, for any combination of SIoT relationships, is evident from the curves shown in Fig. 4; the values in terms of mean IRN percentage obviously increases with the increase in the percentage of nodes that spread the Source VUIP.

G. Performance by varying the type of Interest

Up till now, in our performance evaluation study we have always considered Interest 3. Obviously, the performance figures may depend on the choice of the interest and it is important to understand how the scenario resulting from a change in the interest influences the performance.

Therefore, a first measurement campaign was aimed at analyzing in which scenarios (each characterized by a different Interest) there is a greater number of nodes belonging to the Giant Component, and what is the increase of this component

when passing from the Friendships case to that Enhanced SIoT. The percentage of nodes belonging to the Giant Component is important because it tells us what is the largest subset of nodes that are connected to each other. Here we consider six hops for the diffusion of the Source VUIP and we establish that the percentage of nodes that spread the VUIP of the Source and that authorize access to their PODs at the Source are both 100% at each hop.

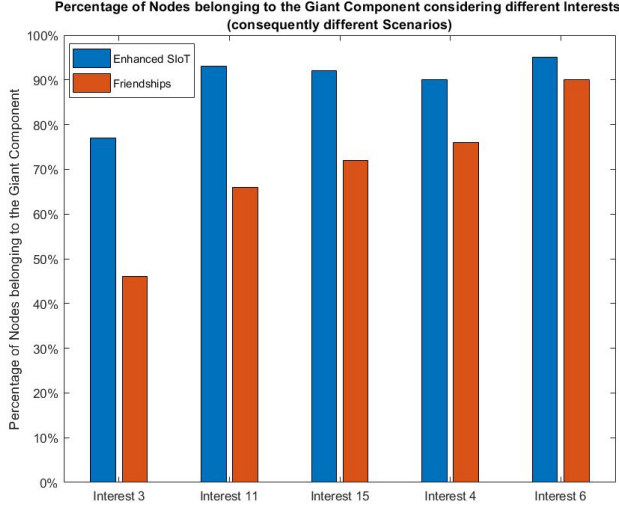


Fig. 5. Percentage of nodes belonging to the Giant Component considering different Interests.

Figure 5 reports the percentage of nodes belonging to the Giant Component, out of the total number of nodes of the Scenario, in the case of Friendships. On the x-axis, the Interests have been sorted by increasing values of nodes belonging to the Giant Component, in the case of Friendships. The percentage of nodes belonging to the Giant Component in the Enhanced SIoT case is also shown in blue.

As the Interest varies, we obtain different scenarios/graphs by considering, for each case, the nodes that possess that specific Interest. Considering only the Friendships as arcs of the graph (Friendships case, red labels) we can obtain graphs in which the nodes are almost all connected to each other (many nodes belong to the Giant Component), as in the Interest 6 case, or in which they are poorly connected, as in the Interest 3 case.

By observing Figure 5 it can be seen that the lower the number of nodes belonging to the Giant Component in the Friendships case, the greater the increase of this value which is obtained in the Enhanced SIoT case (for ease of reading, Figure 6 shows such an increase). This means that the worse the starting scenario (in the case of Friendships), the greater the gain achieved with the Enhanced SIoT. The difference in the results obtained is due to the fact that, as expected, in poorly connected Scenarios in the Friendships case, the established Social Object Relationships are able to connect a larger number of nodes that were not already connected by Friendships.

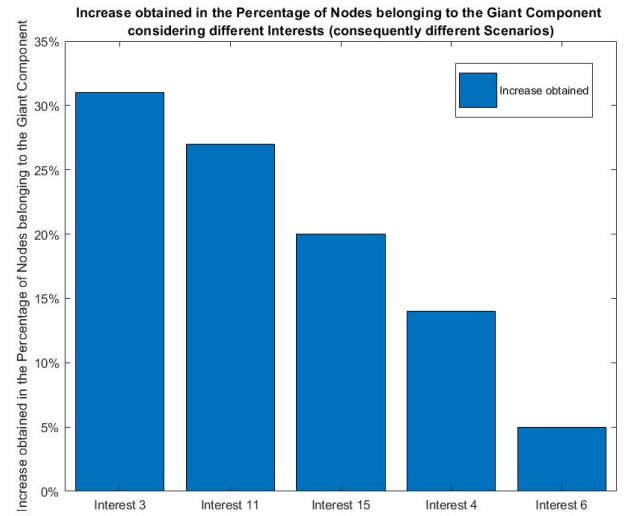


Fig. 6. Increase obtained in the percentage of nodes belonging to the Giant Component considering different Interests.

We need now to better understand which is the increase in the percentage of IRN nodes, which play an important role in our distributed social network. Without losing generality, we focus on Interests 3, 4, and 6, and again consider six hops for the diffusion of the Source VUIP, and 100% of nodes that, at each hop, spread the Source VUIP and authorize access to their PODs.

In Figure 7 we can see that in the graph obtained considering Interest 6, already well connected in the Friendships case, the obtained Mean IRN percentage value is very high. On the contrary considering Interest 3, with only a few nodes belonging to the Giant Component in the Friendships case, the obtained Mean IRN percentage value is very low.

The much more important advantage obtained by using the Social Object Relationships are confirmed in the Interest 3 case, wherein it allows to greatly improve the Mean IRN percentage value. On the contrary, using the Social Object Relationships is of little relevance in the Interest 6 case, wherein the advantage obtained is small since high values of Mean IRN are attainable without its introduction. Like before, if we do not consider the interested but isolated nodes, the behaviours remain the same while the reachable performance levels are higher, as shown in Figure 8

The last study conducted aimed at demonstrating and quantifying the advantage deriving from the adoption of the C-IOR object friendships in terms of mean reduction of the number of hops employed by the Source to reach all the interested nodes during the Discovery phases. A faster discovery enabled by the use of C-IORs has been confirmed, as in our experiments we have always found for each source an average number of hops that is almost halved compared to the case in which C-IORs are not used (an example referring to three nodes randomly chosen is depicted in Figure 9, but a similar behaviour is found for all nodes in the considered population). The manifest

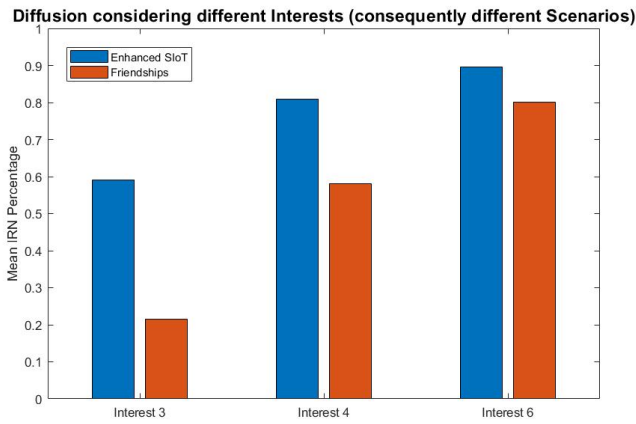


Fig. 7. Mean IRN percentage when varying the considered Interest.

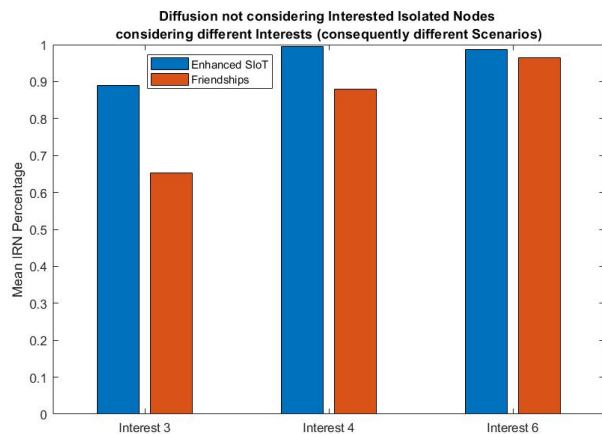


Fig. 8. Mean IRN percentage when considering different Interests (isolated nodes NOT considered).

advantage in terms of delay reduction during Discovery is only paid in terms of a slight increase in the computational complexity introduced by the basic mechanism of the C-IOR establishment.

H. Final remarks

In conclusion, in all the conducted studies the advantage achieved in the Enhanced SIoT case compared to the Friendships case is evident, thanks to the possibility of using the SIoT relationships. Furthermore, the contribution given by the C-IOR relationship appears to be significant. The reason for this is the fact that the basic mechanism of C-IOR allows the establishment of direct social links (C-IOR) to interested nodes that would otherwise be connected only through a chain of SIoT relationships that might also involve nodes not interested to the content.

As for the use of hub nodes as preferential relaying nodes in order to optimize the performance, we have to clarify that for the purposes of our work, the number of links in the social graph of a given node is not so important, but rather the nature

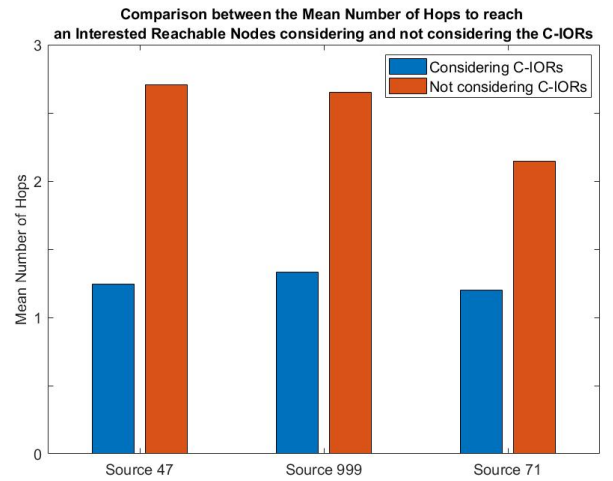


Fig. 9. Comparison between the Mean Number of Hops employed by the Source to reach an Interested Reachable Node when considering/not considering the C-IORs.

of these links (i.e., the nature of the SIoT friendship that the link represents) is important. In fact, a hub node with a very high degree of connection but only with links to other nodes of its own community of interest is much less attractive, for the purposes of its mediator role, than a node that is not a hub and has just a few SIoT links but with objects from other communities with similar interests.

In our work we have assumed the Source S node sending its VUIP to all its first social neighbors (nodes with which S has already established at least one Social Object Relationship) and so on recursively. Furthermore, among the Social Object Relationships it has been noted that POR links (long range) are particularly important to connect different communities that share the same interest. In our simulations, as we are dealing with networks that are not too large in terms of the number of nodes and arcs, we have chosen to consider when possible all the nodes (except in cases with percentages other than 100%).

In real cases, obviously, to avoid spreading to all nodes, relying on hub nodes (intended, however, as nodes with a large number of POR relationships) could certainly bring advantages. This could be a starting point for a later work.

A final point that we intend to highlight relates to possible extensions of the concepts introduced in this paper also to environments other than those of traditional digital social networks, which could be the subject of future investigations. Certainly, a very promising environment in which to test the potential of the new concepts proposed is that of a Vehicular Social Network (VSN). In fact, VSNs are intrinsically characterized by a decentralized nature and present a scenario with cars playing the role of "content prosumers", i.e. both producers and consumers, as it is understood in our work. An aspect that differentiates VSNs from the OSNs we have considered in this work is the *highly dynamic* nature of VSNs, in which social links are built "on-the-fly" and have a short lifetime, whenever the community members become neighbors to each other [52],

[53]. More specifically, a vehicle can enter a social network and stay for a limited time by exchanging messages with the neighboring vehicles about a given topic related to the same interest or experiences, and it can participate to a known social network whenever approaching a specific area of interest [52], [53]. Given these differences, the result of our research cannot be applied “as it is” to VSN environments, since we aim to increase the level of connectivity between communities that belong to a distributed human OSN but which are defined via fairly stable links, through the use of SIoT, which is also decentralized and based on stable inter-device links. Besides, there would be a need to include concepts that are consistent with the destructive paradigm introduced by SOLID also in the VSN environment.

It can be concluded that there is a need for some adaptations to extend what is described in the paper to be applied to vehicular scenarios. There are however good possibilities to achieve this, if we consider that the concept of clustering a VSN into social groups composed of vehicles with common interests is already widely spread, and that the SIoT principles have also been integrated with IoT features in the vehicular environment [54], [55].

VI. CONCLUSIONS

In this paper we have proposed a new platform model for DOSNs based on the joint use of the Solid platform and the new paradigm of SIoT, emerging with increasing strength. Evidence has been provided of the fact that by coupling these two concepts together it is possible to arrive at the design of a modern DOSN platform that permits users to maintain control over their personal information and, at the same time, effectively limits the intrinsic drawbacks that in the past made DOSNs unattractive compared to centralized solutions. Through a simulation campaign aimed at comparing the ability to connect users with the same interest but belonging to separate communities within a DOSN platform, it was possible to prove that the road traced has the potential to make distributed social networks more attractive and to facilitate their large-scale deployment. This can be achieved thanks to the synergies that can be obtained between human users and social devices.

REFERENCES

- [1] Guidi, Barbara, et al. “Managing social contents in decentralized online social networks: a survey.” *Online Social Networks and Media* 7 (2018): 12-29.
- [2] Aiello, Luca Maria, and Giancarlo Ruffo. “Secure and flexible framework for decentralized social network services.” 2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). IEEE, 2010.
- [3] Buchegger, Sonja, et al. “PeerSoN: P2P social networking: early experiences and insights.” *Proceedings of the Second ACM EuroSys Workshop on Social Network Systems*. 2009.
- [4] Cuttillo, Leucio Antonio, Refik Molva, and Thorsten Strufe. “Safebook: A privacy-preserving online social network leveraging on real-life trust.” *IEEE Communications Magazine* 47.12 (2009): 94-101.
- [5] Guidi, Barbara, et al. “DiDuSoNet: A P2P architecture for distributed Dunbar-based social networks.” *Peer-to-Peer Networking and Applications* 9.6 (2016): 1177-1194.
- [6] Yeung, Ching-man Au, et al. “Decentralization: The future of online social networking.” *W3C Workshop on the Future of Social Networking Position Papers*. Vol. 2. 2009.
- [7] Diaspora. <https://diasporafoundation.org/>.
- [8] Samba, Andrei Vlad, et al. “Solid: a platform for decentralized social applications based on linked data.” Technical report, MIT CSAIL & Qatar Computing Research Institute, 2016.
- [9] Mansour, Essam, et al. “A demonstration of the solid platform for social web applications.” *Proceedings of the 25th International Conference Companion on World Wide Web*. 2016.
- [10] Gao, Wei, and Guohong Cao. “User-centric data dissemination in disruption tolerant networks.” 2011 *Proceedings IEEE INFOCOM*. IEEE, 2011.
- [11] Mashhadi, Afra J., Sonia Ben Mokhtar, and Licia Capra. “Habit: Leveraging human mobility and social network for efficient content dissemination in delay tolerant networks.” 2009 *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks & Workshops*. IEEE, 2009.
- [12] Kong, Chenguang, and Xiaojun Cao. “Semi-controlled authorized information dissemination in content-based social networks.” 2014 23rd *International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2014.
- [13] Kong, Chenguang. “Content Dissemination in Mobile Social Networks.” (2016).
- [14] Karin Knorr Cetina, chapter on ‘Objectual Practice’ in “The practice turn in contemporary theory”, Karin Knorr CETINA, Theodore R. SCHATZKI, Eike VON SAVIGNY (ed.), Routledge, 2005.
- [15] Atzori, Luigi, et al. “The social internet of things (siot)—when social networks meet the internet of things: Concept, architecture and network characterization.” *Computer networks* 56.16 (2012): 3594-3608.
- [16] Atzori, Luigi, Antonio Iera, and Giacomo Morabito. “From “smart objects” to “social objects”: The next evolutionary step of the internet of things.” *IEEE Communications Magazine* 52.1 (2014): 97-105.
- [17] Chen Ray, Fenyao Bao, and Jia Guo. “Trust-based service management for social internet of things systems.” *IEEE transactions on dependable and secure computing* 13.6 (2016): 684-696.
- [18] Chen Zhikui et al. “A scheme of access service recommendation for the Social Internet of Things.” *Int. J. Communication Systems* 29 (2016): 694-706.
- [19] Nitti Michele, et al. “A subjective model for trustworthiness evaluation in the social internet of things.” *Personal Indoor and Mobile Radio Communications (PIMRC)*, 2012 *IEEE 23rd International Symposium on*. IEEE, 2012.
- [20] Lin, Zhiting, and Liang Dong. “Clarifying trust in social internet of things.” *IEEE Transactions on Knowledge and Data Engineering* 30.2 (2017): 234-248.
- [21] Atzori, Luigi, Antonio Iera, and Giacomo Morabito. “Understanding the Internet of Things: definition, potentials, and societal role of a fast evolving paradigm”, *Ad Hoc Networks*, 56 (2017): 122-140.
- [22] Schurgot, Mary R., Cristina Comaniciu, and Katia Jaffres-Runser. “Beyond traditional DTN routing: social networks for opportunistic communication.” *IEEE Communications Magazine* 50.7 (2012): 155-162.
- [23] Atzori, L., Campolo, C., Da, B., Girau, R., Iera, A., Morabito, G., & Quattropani, S. (2019), “Smart Devices in the Social Loops: Criteria and Algorithms for the Creation of the Social Links”, *Future Generation Computer Systems*, Volume 97, August 2019, Pages 327-339
- [24] Guidi, Barbara, Andrea Michienzi, and Giulio Rossetti. “Towards the dynamic community discovery in decentralized online social networks.” *Journal of Grid Computing* 17.1 (2019): 23-44
- [25] Mega, Giuliano, Alberto Montresor, and Gian Pietro Picco. “Social Overlays Meet the Cloud: A Hybrid Architecture for Profile Dissemination in Decentralized Social Networks.” *IEEE Transactions on Network Science and Engineering* 6.4 (2018): 613-627.
- [26] Conti, M., De Salve, A., Guidi, B., and Ricci, L. (2014). Epidemic diffusion of social updates in dunbar-based dosn. In *European Conference on Parallel Processing*, pages 311-322. Springer.
- [27] Yuan, Bo, Lu Liu, and Nick Antonopoulos. “Efficient service discovery in decentralized online social networks.” *Future Generation Computer Systems* 86 (2018): 775-791.
- [28] Paul, Thomas, Sonja Buchegger, and Thorsten Strufe. “Decentralized social networking services.” *Trustworthy Internet*. Springer, Milano, 2011. 187-199.

- [29] Buchegger, S. and Datta, A. (2009). A case for p2p infrastructure for social networks opportunities & challenges. In 2009 Sixth International Conference on Wireless On-Demand Network Systems and Services, pages 161–168. IEEE.
- [30] Graffi, K. and Masinde, N. (2020). Libresocial: A peer-to-peer framework for online social networks. arXiv preprint arXiv:2001.02962.
- [31] Guidi, Barbara, et al. "The Contextual Ego Network P2P Overlay for the Next Generation Social Networks." *Mobile Networks and Applications* (2020): 1-13.
- [32] Mastodon. <https://mastodon.social/about>.
- [33] Li, Zechao, Jinhui Tang, and Tao Mei. "Deep collaborative embedding for social image understanding." *IEEE transactions on pattern analysis and machine intelligence* 41.9 (2018): 2070-2083.
- [34] Nordström, Erik, Christian Rohner, and Per Gunningberg. "Haggle: Opportunistic mobile content sharing using search." *Computer Communications* 48 (2014): 121-132.
- [35] Wang, Zhixiao, Li, Zechao, et al. "Tracking the evolution of overlapping communities in dynamic social networks." *Knowledge-Based Systems* 157 (2018): 81-97.
- [36] Birrane, Ed and Soloff, Jason. "Designing Delay-Tolerant Applications for Store-and-Forward Networks", Artech House, 2020.
- [37] Boldrini, Chiara, Marco Conti, and Andrea Passarella. "ContentPlace: social-aware data dissemination in opportunistic networks." *Proceedings of the 11th international symposium on Modeling, analysis and simulation of wireless and mobile systems*. 2008.
- [38] Ciobanu, Radu-Ioan, et al. "Onside: Socially-aware and interest-based dissemination in opportunistic networks." 2014 IEEE Network Operations and Management Symposium (NOMS). IEEE, 2014.
- [39] Mega, Giuliano, Alberto Montresor, and Gian Pietro Picco. "On churn and communication delays in social overlays." 2012 IEEE 12th International Conference on Peer-to-Peer Computing (P2P). IEEE, 2012.
- [40] Zhou, Jiang, Rami Albatat, and Cathal Gurrin. "Applying visual user interest profiles for recommendation and personalisation." *International Conference on Multimedia Modeling*. Springer, Cham, 2016.
- [41] Girau, Roberto, Salvatore Martis, and Luigi Atzori. "Lysis: A platform for IoT distributed applications over socially connected objects." *IEEE Internet of Things Journal* 4.1 (2016): 40-51.
- [42] Njoo, Gunarto Sindoro, Kuo-Wei Hsu, and Wen-Chih Peng. "Distinguishing friends from strangers in location-based social networks using co-location." *Pervasive and Mobile Computing* 50 (2018): 114-123.
- [43] Cho, Eunjoon, Seth A. Myers, and Jure Leskovec. "Friendship and mobility: user movement in location-based social networks." *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2011.
- [44] Marche, Claudio, Luigi Atzori, and Michele Nitti. "A dataset for performance analysis of the social internet of things." 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). IEEE, 2018.
- [45] G. W. Index. "GWI Device Q1 2017." Available from <https://cdn2.hubspot.net/hubfs/304927/Downloads/GWI-Device-Q1-2017-Summary.pdf>, 2017.
- [46] Scellato, Salvatore, et al. "Socio-spatial properties of online location-based social networks." *Fifth international AAAI conference on weblogs and social media*. 2011.
- [47] Cheng, Zhiyuan, et al. "Exploring millions of footprints in location sharing services." *Fifth International AAAI Conference on Weblogs and Social Media*. 2011.
- [48] Yang, Dingqi, Daqing Zhang, and Bingqing Qu. "Participatory cultural mapping based on collective behavior data in location-based social networks." *ACM Transactions on Intelligent Systems and Technology (TIST)* 7.3 (2016): 1-23.
- [49] Yang, Dingqi, et al. "Nantotelescope: Monitoring and visualizing large-scale collective behavior in lbsns." *Journal of Network and Computer Applications* 55 (2015): 170-180.
- [50] Yang, Dingqi, et al. "Revisiting user mobility and social relationships in lbsns: A hypergraph embedding approach." *The World Wide Web Conference*. 2019.
- [51] Marche, Claudio, et al. "Navigability in social networks of objects: The importance of friendship type and nodes' distance." 2017 IEEE Globecom Workshops (GC Wkshps). IEEE, 2017.
- [52] Vegni, Anna Maria, and Valeria Loscri. "A survey on vehicular social networks." *IEEE Communications Surveys & Tutorials* 17.4 (2015): 2397-2419.
- [53] Rahim, A., Kong, X., Xia, F., Ning, Z., Ullah, N., Wang, J., Das, S. K. (2018). Vehicular social networks: A survey. *Pervasive and Mobile Computing*, 43, 96-113.
- [54] K. Alam, M. Saini, and A. El Saddik, "tNote: A social network of vehicles under internet of things," in *Internet of Vehicles Technologies and Services*, vol. 8662, ser. Lecture Notes in Computer Science, R.-H. Hsu and S. Wang, Eds. New York, NY, USA: Springer-Verlag, 2014, pp. 227–236.
- [55] M. Nitti, R. Girau, A. Floris, and L. Atzori, "On adding the social dimension to the Internet of vehicles: Friendship and middleware," in *Proc. IEEE Int. BlackSeaCom Netw.*, May 2014, pp. 134–138.