REGULAR PAPER

Incremental communication patterns in online social groups

Andrea Michienzi[1] · Barbara Guidi[1] · Laura Ricci[1] · Andrea De Salve[2]

Abstract

In the last decades, temporal networks played a key role in modelling, understanding, and analysing the properties of dynamic systems where individuals and events vary in time. Of paramount importance is the representation and the analysis of Social Media, in particular Social Networks and Online Communities, through temporal networks, due to their intrinsic dynamism (social ties, online/offline status, users' interactions, etc..). The identification of recurrent patterns in Online Communities, and in detail in Online Social Groups, is an important challenge which can reveal information concerning the structure of the social network, but also patterns of interactions, trending topics, and so on. Different works have already investigated the pattern detection in several scenarios by focusing mainly on identifying the occurrences of fixed and well known motifs (mostly, triads) or more flexible subgraphs. In this paper, we present the concept on the Incremental Communication Patterns, which is something in-between motifs, from which they inherit the meaningfulness of the identified structure, and subgraph, from which they inherit the possibility to be extended as needed. We formally define the Incremental Communication Patterns and exploit them to investigate the interaction patterns occurring in a real dataset consisting of 17 Online Social Groups taken from the list of Facebook groups. The results regarding our experimental analysis uncover interesting aspects of interactions patterns occurring in social groups and reveal that Incremental Communication Patterns are able to capture roles of the users within the groups.

Keywords Temporal networks · Online social networks · Online social groups · Communication patterns

✉ Andrea Michienzi
andrea.michienzi@di.unipi.it

Barbara Guidi
guidi@di.unipi.it

Laura Ricci
laura.ricci@unipi.it

Andrea De Salve
andrea.desalve@unipa.it

1 Department of Computer Science, University of Pisa, Pisa, Italy

2 Department Of Mathematics and Computer Science, University of Palermo, Palermo, Italy

1 Introduction

Online Social Networks [14] (OSNs) are cornerstones of today's communication. Indeed, OSNs such as Facebook are nowadays used as the main communication channel by people because of their power and versatility [27]. It is very common to see people share their personal information, photos, etc., with friends and with strangers [15]. This communication mechanism is particularly relevant in virtual online communities, which represent one of the most important emerging feature of OSNs. Indeed, a current trend of Social Media is to offer members the opportunity to establish and join groups of people online by creating virtual communities based on similar interests. These virtual communities can be called Online Social Groups (OSGs) [36,40], and they model a set of users interacting in discussions about real-world events, hobbies, similar interests, etc. In OSGs people have the chance to discuss ideas, doubts or interests with a big number of people that would be otherwise difficultly reachable in real life or with other traditional communication systems, that are mostly dedicated to private, one-to-one ways of messaging [10,12]. These OSGs create environments where the information spreading tend to be even more effective [4,13,24] than the one seen in the word-of-mouth effect [1]. And, even more than that, in contrast to how an OSN is usually used, where a user knows all of its contacts, users in OSGs do not necessarily know each other, and there is no way of limiting the interactions within the group.

An OSG is defined in [2] as a collection of people which can be divided into two categories: an extension of social identification (individuals affiliate with organizational memberships, gender, age, etc.) and related to structured communication, built around communication (social support, political debate, or similar interests). Despite the growing importance of OSNs in facilitating social communication, there has been limited research focusing on the communication patterns in OSGs. OSGs, as typical complex networks, can be modelled by graphs which are time-dependent [6]. Indeed, they need to be studied as dynamic networks to understand their real characteristics, as demonstrated in [16,17] for what concerns the study of community detection.

1.1 Motivations and contributions

OSGs can be represented as temporal graphs, and significant recurring patterns of interaction, or communication patterns, between nodes can be found. Communication patterns can be identified using several graph concepts: from motifs [28,29] to graphlets [23], from subgraphs [37] to temporal greedy walks [38]. In particular, the study of temporal motifs has attracted a lot of interest and showed how motifs can be helpful to understand particular characteristics of the human behaviour, such as homophily [9,42], mobility [39], preferences [35], analysis of trends [26], but also human brain [3,11], stock prediction [25], weather prediction [31] and many other fields. Today, one of the main problems to understand communication patterns with current proposals is the use of fixed "size" structures [23]. Especially in the case of motif detection, that is the most used approach, they only deal with structures of limited size in terms of nodes and edges involved in the motif due to the complexity of the problem. This introduces a very big limitation of the usefulness of the approach if we want to capture arbitrarily complex communication patterns. The study of small and fixed motifs can help in the understanding of common communication patterns, but the communication patterns do not have, in principle, a fixed structure and are not limited to few nodes/edges. Other notions, such as the one of subgraphs and temporal greedy walks, try to overcome this limitation by building models that are not limited in this direction. However, they completely lose any

kind of causality relation between the interactions that make up the pattern by not fixing any structure. In this paper, we propose a new model for the identification of patterns in OSGs that overcome the constraint of the limited size proper of the motifs and graphlets, and, at the same time, overcome the lack of structure and explicit causality relation of the temporal subgraphs and temporal greedy walk. To this aim, we propose and define the concept of Incremental Communication Pattern. We apply this concept to a set of Facebook Groups, because the study of temporal communication patterns in the scenario of OSGs is still missing. We study a set of five patterns designed specifically for social networks, such that possible specific roles of the users within the OSGs emerge. For the sake of readiness, we propose the following contributions:

- we propose and formalize the definition of Incremental Communication Pattern, using a generic temporal graph formalism, by the means of a basic pattern and an incremental rule;
- we propose and formalize five Incremental Communication Patterns which are crucial in OSGs to identify specific communication patterns and the social role of actors involved in;
- we study both the recurring and the maximum size, defined as the number of edges appearing in the largest pattern detected, for each Incremental Communication Pattern, in order to understand the complexity of the related communication, by exploiting a real dataset of 17 Facebook groups.

The rest of the paper is structured as follows. In Sect. 2, we present an overview of the state of the art in terms of motifs, graphlets, subgraphs and greedy walks in OSNs. In Sect. 3, we describe in detail our scenario, and in Sect. 4 we propose our idea of Incremental Communication Patterns. In Sects. 5 and 6, we describe the dataset and the obtained results, respectively. Finally, in Sect. 7 we draw the conclusions and propose some possible future works.

2 Related work

Many complex networks, such as Online Social Networks, are considered temporal networks [22]. Indeed, relationships appear and disappear due to the temporal evolution of the social relationships [20], or due to the offline/online state of users. Usually, social networks are modelled by using a graph. Graphs are a natural representation of a set of entities and the relationships among them, and by considering a social network as a temporal network, temporal graphs are the general model used to model their dynamic nature. Small subgraph patterns, such as motifs or graphlets, together with other indicators, are crucial to understand the structure and the evolution of the graphs [34]. When dealing with social networks as temporal networks, patterns are affected by the temporal order of interactions, and it is represented by a temporal motif, which take into account the changes of the network. A first notion of a temporal network motif, proposed in [34], define it as sequences of edges that are time-ordered and confined within a temporal interval of length. Instead, in [29] temporal motifs are defined as equivalence classes of temporal subgraphs whose events happen at a distance smaller than a fixed value and are consecutive. For what concerns the studies oriented to evaluate the temporal motifs in complex networks, important contributions have been presented in the field of communication networks (mobile or social networks), because temporal motifs are a useful tool to represent and analyse temporal communication patterns in

order to find important characteristics of a social network, but also anomalies and unexpected patterns.

Zhao et al. [47] extrapolate motifs from two datasets derived from Call Detail Records and Facebook. Authors find the most common motifs using a time window of 4 hours. In [46], authors explore 3-event temporal motifs in six datasets to understand the human interactive patterns, and discovered three dominant temporal motifs: Star, Ordered-chain, and Ping-Pong, with the corresponding interactive patterns of Leader, Queue, and Feedback, respectively. Kovanen et al. [29] take advantage of the null model concept to show how factors like sex and age may have an influence on communication patterns. This property is called temporal homophily, i.e. the tendency of individuals with the same attributes to participate in the same communication patterns. Creusefond et al. [7] study temporal motifs by using the communities structure. From the experiments, they observe star and chain motifs are frequent inside communities where there are the same set of actors, instead spams, ping-pong, and triangles occur in different communities (inter-community edges). In [5], authors find triadic motifs in a Science Library dataset, because a triad is a pattern that connect three actors, and it is considered to be fundamental structural pattern of social networks [21]. In [48], authors focus on cohesive social groups built by exploiting relationships by mobile phone. They propose a methodology to identify cohesive groups and extract temporal motifs to show how members of social groups interact by means of calls and text messages. Liu et al. [30] analyse temporal motifs by using the notion of stochastic temporal network motif which models the sequential dependency of communications within a communication pattern using a first-order Markov chain. Xuan et al. [45] use temporal motifs to reveal collaboration patterns in task-oriented social networks, which are networks contain different types of nodes and links and identify people which collaborate to produce different kinds of artifacts, such as movies, music, etc. In [32], the author uses the concept of temporal motif to characterize the behaviour of the nodes in the network. Lastly, Wu et al. [44] enhance the definition of temporal motifs considering the labels on the edges.

Also graphlets have been studied in a temporal fashion. In [23], authors tackle the problem of counting all the possible graphlets with a given number of nodes and edges. They also try to characterize graphs and nodes of the graphs based on the graphlets found. Even though they present an approach which is not limited in the number of nodes or edges involved in the graphlet, they limit the experimental evaluation to graphlets made of up to 10 nodes and 10 edges.

The idea of Time-respecting subgraph, introduced in [37], is very similar to the one of graphlets. The idea is to put together all edges which share at least one end and that happen within a threshold. The idea is based on executing a breadth-first search starting from a given seed interaction, but adding a temporal constraint to the problem.

Finally, authors of [38] study the greedy walks on temporal networks to detect the burst trains and dominant factors. The intuition behind this work is that, if a set of users is particularly active and interact a lot each other, a temporal greedy walk will remain struck among these few nodes. Also this approach is not limited in the number of nodes and edges that can be included in the structure; however, the returned walk may not be relevant because there is no needed structure and because of the greediness of the walks returned.

At the best of our knowledge, what is lacking in literature is a study which focuses on the longest meaningful patterns that one can find. While they give generic definitions, current proposals of temporal motifs only study the problem with a fixed and rather low length, which means a few number of nodes (generally speaking no more than 4). Other works still focus on short patterns and count the most frequent ones, without really grasping the fact that human communication is not bound in size, or find structureless patterns, that are not capable

to isolate very specific and out-of-the box behaviours. Instead, in this work we define and analyse Incremental Communication Patterns in order to discover recurring and structured sets of interactions, with a particular interest in the maximum pattern length to understand specific social roles, such as influencers or bots.

3 Foundamentals

In this Section, we introduce the reader to the Online Social Groups, that is the scenario we consider in this paper, and to the basics of the modelling of interactions via a temporal graph.

3.1 Online social groups

One of the main current functionality offered by Social Media is the possibility of creating groups of Social Media members [10,43], such as Facebook Groups, the circles in Google+, etc. A specific property of OSGs is that people do not necessarily need to know all the other members of the group, meaning that no explicit consent is required to start interacting.

Inside OSGs, users can typically write posts, and other users can interact with posts in a number of ways. The most significant way of interaction with posts is via written comments [12]. In many OSGs platforms, users can also interact with these comments with other comments, just like with the post. This creates a potentially infinite tree-like structure of interactions between users which enables very complex communication structures among them. In this work, we focus on such OSG scenario because of the ever increasing interest of people to join virtual communities and the lack of a depth analysis of both the characteristics and the communication patterns in this specific case. Indeed, the study of the temporal communication patterns is very important to understand the role of a user in a virtual community, such as a social group. Roles of the users, such as the so-called influencers, are usually analysed by exploiting centrality measures on a graph representing all the interactions of the users. However, this strategy is too much general and do not take into account a causality relation among the interactions of the users. Thanks to the idea of Incremental Communication Patterns, we can not only identify the most frequent communication patterns, but also study the characteristics of the communication (engagement, length, etc.) to detect the activity of users in a time span. This is a tool to study anomalies and specific social roles of users because the rules that defines them try to capture specific behaviours of the OSG scenario. Finally, it is also important to note that the proposed approach is also valid in the case of OSNs analysis, and it can be also used in several other contexts, such as analysis of network packets, of bitcoin transaction network, and of email communication networks to name a few.

3.2 Temporal multigraph: definition

To study communication patterns in OSGs, we are interested in modelling how the users in the group interact with each other. To this aim, we introduce the interaction graph as a graph in which the nodes represent the users of the OSG and the edges represent the interactions between users. It is also of primary importance to enrich the graph with labels on the edges of this graph corresponding to the timestamp at which the interaction happened in order to establish possible causality relations. Such interaction graph can be modelled with a temporal multigraph $G = (V, E)$. A temporal multigraph $G = (V, E)$ is a graph where V is the set of the nodes in the graph, E is the set of edges and each edge $e \in E$ is a tuple $e = (s, d, t)$,
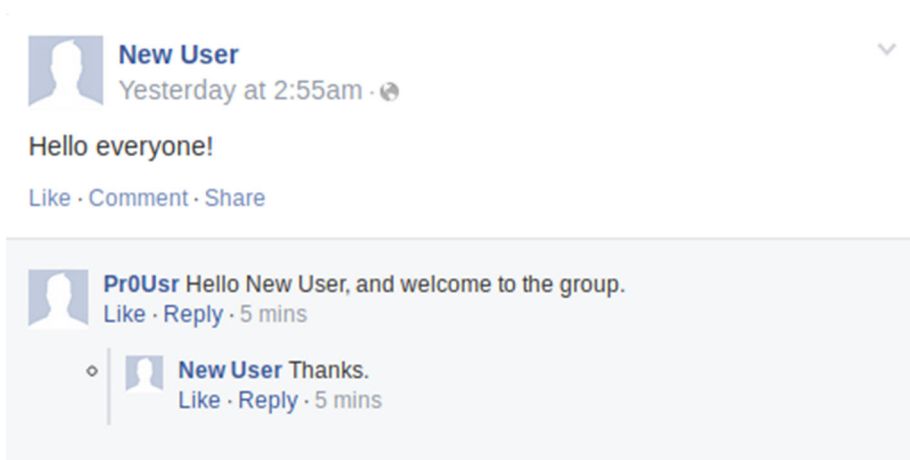
Fig. 1 A fake post of an OSG. Considering this interaction, we will record two edges. The first edge is (Pr0Usr, New User, t$_1$) and corresponds to the comment by Pr0Usr to the post made by New User, the second one is (New User, Pr0Usr, t$_2$) and corresponds to the comment by New User to Pr0Usr

where $s \in V$ is the source of the edge, $d \in V$ is the destination of the edge, and $t \in \mathbb{R}$ is the timestamp at which an event, an interaction in our scenario, happens. Using the temporal multigraph formalism, we can model the users as the set of nodes V of the graph, and the interactions as the set of edges E. As we already said in Sect. 3.1, we identify as interactions comments to posts and comments to other comments, but we do not consider writing a post an interaction itself. While it easy to know who wrote the post, it is not easy to determine to whom in particular the post is addressed to. For this reason, we assume that posts are just handles from which real interactions can start. For each comment to a post, we will create an edge with source the node who wrote the comment and, as destination, the author of the post. Note that if a user writes a comment to one of its post, an interaction towards itself is generated. Analogous is the case in which a user writes a reply to a comment: In this case, an interaction is generated from the author of the reply to the author of the comment. For example, consider the case in which user A writes a post at time t_a, then user B writes a comment to A's post at time t_b, and finally user C writes a comment to B's comment (see Fig. 1 for an example). This results in two edges in the graph: (B, A, t_b) and (C, B, t_c). Due to the possible presence of bots and other tools for spamming, we may encounter two interactions originating from the same user, directed to the same, happened at the same time. To be able to distinguish these two interactions, we suppose that each interaction has an unique identifier.

4 Incremental communication patterns

In this section, we introduce the concept of Incremental Communication Pattern giving a generic idea of its structure and the motivation behind it. Then we give the definitions of 5 social patterns specifically thought for the scenario of OSGs. A communication pattern can be simply seen as a set of interactions which underlines that a specific communication happened among a set of users. The communication pattern does not only specify the direction of each interaction, but it also specifies the relations among the involved users, because the

Fig. 2 The disadvantage of having fixed time windows. **a** a fixed time window approach may cause interactions belonging to the same pattern, denoted with "X" in the figure, to be split in two distinct time windows. **b** the timeout approach is much better for searching maximal patterns
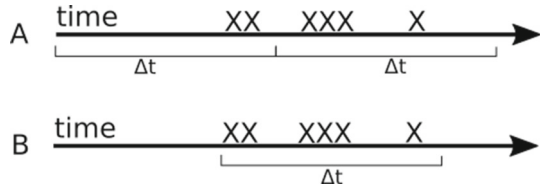
interactions have explicit sources and destinations. The biggest limitations of (fixed) communication patterns are that they are not able to capture, per se, the fact that communication is not bound by the number of users involved or by the number of interactions. To overcome this limitation, we introduce the concept of Incremental Communication Pattern. An Incremental Communication Pattern enhances the general idea of a communication pattern, usually composed by few users and interactions, by introducing a rule to add more users and interactions to the pattern. By applying recursively the rule, we can thus generate more complex communication patterns in an incremental fashion.

To detect an Incremental Communication Pattern, it is necessary to find, in the temporal graph of interactions, a basic pattern and then, to apply as many times as possible the incremental rule to create the largest instance of the communication pattern, according to the order in which the interactions appear. To be able to classify the different instances of the patterns, which depend on how many times we were able to apply the incremental rule to a basic pattern, we also define the *size* of an incremental pattern as the number of edges appearing in the pattern. We will use the letter k to address to the size of the pattern and say that it has size k if a specific instance of the pattern contains k temporal edges. As we will present later in this section, the basic rule defines small patterns of size 2. It is impossible to define a smaller basic pattern for two reasons: A smaller basic pattern would cause confusion as the basic patterns cannot be distinguished from each other; moreover, a pattern consisting of one edge is just an interaction, and it completely loses any structure. It is also worthwhile to notice that there is not any upper bound of the size of an instance of an Incremental Communication Pattern, due to its recursive definition.

The idea we propose is, to a certain degree, closely related to the one of motif, already present in literature. In fact, finding an occurrence of the basic pattern in the graph is the same problem of motif detection, that is a subgraph isomorphism problem. However, we model the evolution of a basic pattern over time due to the addition of arbitrary interactions, we use an incremental rule, instead of defining a new motif.

Another very important aspect to consider when detecting such communication patterns, without a fixed structure, is the time. In particular, we want to set a *timeout* for the pattern, that is an amount of time within all the interactions of a given instance of a pattern have to appear to be considered part of the pattern. A similar idea is already present in literature and it is called time window [41], but the concept of time window is usually bound to the absolute time dimension. While this is still a reasonable approach in many fields, this is not the case in our scenario. In fact, using a fixed time window, that sums up in slicing time in fixed size buckets and then study the problem within each bucket separately, may introduce a problem. Suppose we have the situation depicted in Fig. 2 and that we choose a time window Δt such that the time is divided as in the case A. If all the interactions are part of the same pattern, we are losing part of the pattern, because of the unfortunate division of time. A better approach would be the one in Fig. 2b, in which we set the start of the detection process the very first interaction of a pattern. In this case, the temporal aspect of the problem is directly bound to each instance of a pattern, and, if the *timeout* Δt is big enough, we are able to

detect the whole pattern, rather than just a part of it. Anyways, the definition of Incremental Communication Pattern is decoupled from the concept of time window, such that a custom time window can be defined depending on the scenario we want to analyse.

In this work, we study and evaluate four specific Incremental Communication Patterns: *k-chain*, *k-in-star*, *k-out-star*, and *k-ping-pong*. Moreover, we introduce a new communication pattern for OSGs, namely the *k-one-way couple*. Each Incremental Communication Pattern presented in this paper expresses an interaction template which can be commonly observed in the context of OSGs, as we will explain in detail in the next subsections. In the following, we firstly give a definition of the Incremental Communication Pattern, in terms of the basic pattern and the incremental rule, and then we discuss the rationale behind the proposed patterns.

4.1 The chain pattern

Definition 1 *(k-chain)* The k-chain pattern can be expressed using the temporal graph formalism with the following basic pattern and incremental rule:

$$
\begin{aligned}
\text{basic pattern}: KC(2) &= (V_{KC_2}, E_{KC_2}) \\
V_{KC_2} &= \left\{ u_i | i \leq 3 \wedge i \in \mathbb{N} \wedge \forall j < i . u_j \neq u_i \right\} \\
E_{KC_2} &= \{(u_2, u_1, t_1), (u_3, u_2, t_2) | t_1 < t_2\} \\
\\
\text{incremental rule}: KC(k+1) &= (V_{KC_{k+1}}, E_{KC_{k+1}}) \\
V_{KC_{k+1}} &= V_{KC_k} \cup \{u_{k+2}\} \\
E_{KC_{k+1}} &= E_{KC_k} \cup \{(u_{k+2}, u_{k+1}, t_{k+1}) | t_k < t_{k+1}\}
\end{aligned}
\tag{1}
$$

The *k-chain* basic pattern consists of a 4 nodes path graph, where each node is linked to the previous and the following one according to their labels. Moreover, it is also required that the interaction happens in a specific order, that is the i-th edge must have, as source, the i-th node, and as destination, the $(i-1)$-th node. The incremental rule of the pattern aims at making the chain longer, trying to attach a new edge, and a new node, after the last node that appeared in the pattern (see Fig. 3).

In our scenario, this motif is important because it symbolizes how much a content is able to engage more and more different users. For instance, if a Facebook post creates a lengthy k-chain pattern, we can say that the original post, and the discussion around it, is extremely engaging. Such pattern can reveal the importance of a content.

4.2 The star patterns

The star is an important pattern which can reveal users that perform particular roles such as the influencer or the bot. In this work, we identify two types of star patterns: the in-star and the out-star. In both cases, at least four nodes connected in a star fashion are required, one of which is the central node and the others are peripheral ones.
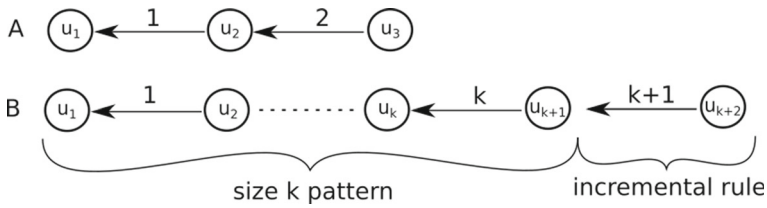
Fig. 3 The k-chain pattern. **a** shows the basic pattern, while **b** shows how the incremental rule increases the size of the pattern

Definition 2 *(k-in-star)* A k-in-star can be expressed using the temporal graph formalism with the following basic pattern and incremental rule:

$$\text{basic pattern}: KIS(2) = (V_{KIS_2}, E_{KIS_2})$$
$$V_{KIS_2} = \{u_c\} \cup \{u_i | i \leq 2 \wedge i \in \mathbb{N}\}$$
$$E_{KIS_2} = \{(u_i, u_c, t_i) | i \in \mathbb{N} \wedge i \leq 2 \wedge t_i < t_{i+1}\}$$

(2)

$$\text{incremental rule}: KIS(k+1) = (V_{KIS_{k+1}}, E_{KIS_{k+1}})$$
$$V_{KIS_{k+1}} = V_{KIS_k} \cup \{u_{k+1}\}$$
$$E_{KIS_{k+1}} = E_{KIS_k} \cup \{(u_{k+1}, u_c, t_{k+1}) | t_k < t_{k+1}\}$$

The basic rule defines a star graph in which all nodes have one outgoing edge, except the central node which has no outgoing edges.

The destination node of all the edges is the central node, which is the node having a special role to be identified. The incremental rule models the ability of the central node to attract even more different interactions towards it (Fig. 4). In fact, a new interaction can be added to the pattern if it involves a new user, and if the interaction has as source node the new user and as destination node the central user.

The communication patterns among members of an OSGs can provide important information about the role played by some individuals of the group. Indeed, a k-in-star pattern represents a communication pattern where the central node attracts a high number of interactions from the other members, resulting in an influence process. In our scenario, the number k of edges of the communication pattern determines the amount of group's members who interacts with the specific user of the group, which is a typical trait of influencers.

Definition 3 *(k-out-star)* A k-out-star can be expressed using the temporal graph formalism with the following basic pattern and incremental rule:

$$\text{basic pattern}: KOS(2) = (V_{KOS_2}, E_{KOS_2})$$
$$V_{KOS_2} = \{u_c\} \cup \{u_i | i \leq 2 \wedge i \in \mathbb{N}\}$$
$$E_{KOS_2} = \{(u_c, u_i, t_i) | i \in \mathbb{N} \wedge i \leq 2 \wedge t_i < t_{i+1}\}$$

(3)

$$\text{incremental rule}: KOS(k+1) = (V_{KOS_{k+1}}, E_{KOS_{k+1}})$$
$$V_{KOS_{k+1}} = V_{KOS_k} \cup \{u_{k+1}\}$$
$$E_{KOS_{k+1}} = E_{KOS_k} \cup \{(u_c, u_{k+1}, t_{k+1}) | t_k < t_{k+1}\}$$

The basic pattern of a k-out-star defines once again a star graph, but this time the direction of the edges is opposite: all edges originate from the central node and have as destination
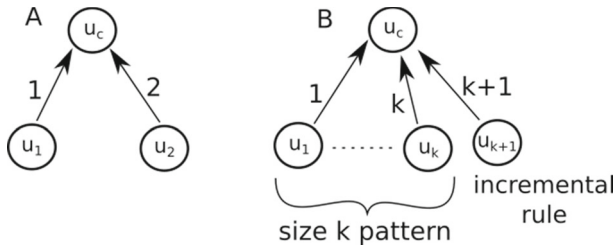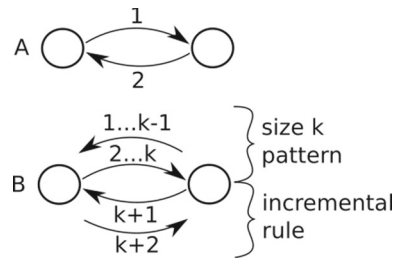
Fig. 4 The k-in-star pattern. **a** shows the basic pattern, while **b** shows how the incremental rule increases the size of the pattern. The k-out-star pattern has the same structure, but source and destination nodes of the edges is inverted

Fig. 5 The k-ping-pong pattern. **a** shows the basic pattern, while **b** shows how the incremental rule increases the size of the pattern

a peripheral node. The incremental rule accepts a new edge with the central node as source node and a new peripheral node as destination node, and it is used to detect new interactions made by the central node towards new users.

In our scenario, this communication pattern is most useful to detect very active users which tend to interact with everyone, and combining this information with other patterns can be useful in detecting spammers and bots. Indeed, we expect these nodes to produce a huge amount of interactions but not receiving any because they are ignored by other real users. We also expect that these users aim to reach a huge number of other users, but not to establish any well-structured communication.

4.3 The ping-pong

Definition 4 *(k-ping-pong)* The k-ping-pong pattern can be expressed using the temporal graph formalism with the following basic pattern and incremental rule:

$$
\begin{aligned}
\text{basic pattern}: KPP(2) &= (V_{KPP_2}, E_{KPP_2})\\
V_{KPP_2} &= \{u, v\}\\
E_{KPP_2} &= \{(u, v, t_1), (v, u, t_2)|t_1 < t_2\}
\end{aligned}
\tag{4}
$$

$$
\begin{aligned}
\text{incremental rule}: KPP(k+2) &= (V_{KPP_{k+2}}, E_{KPP_{k+2}})\\
V_{KPP_{k+2}} &= V_{KPP_k}\\
E_{KPP_{k+2}} &= E_{KPP_k} \cup \{(u, v, t_{k+1}), (v, u, t_{k+2})|t_k < t_{k+1} < t_{k+2}\}
\end{aligned}
$$

This pattern is possibly the most particular one as the basic pattern is defined for $k = 2$, and thanks to the incremental rule it is defined only for every even k. The basic pattern of

the k-ping-pong is composed by two nodes u and v and two edges connecting them. One edge has as source node u and as destination node v, while the other edge has as source node v and as destination node u. The incremental rule aims at detecting new occurrences of an interaction happening from u to v directly followed by another interaction happening in the opposite direction. In fact, the incremental rule adds two edges to the pattern, and this is why the pattern is defined only for even k (see Fig. 5).

In our scenario, this pattern is mostly useful to detect pair of people which tend to interact a lot in a riposte and counter-riposte fashion. Moreover, each edge making the sequence must have inverted source and destination node with respect to the previous one in the sequence. Therefore, we aim to capture a causality relation between interactions happening between the two users. The k-ping-pong pattern allows us to recognize users who are involved in reciprocal interactions (such as discussion among members which goes on as long as one reply the other).

4.4 The one way couple pattern

This pattern is not present in literature, and it is part of our contribution, although it is similar to the k-ping-pong.

Definition 5 (*The k-one-way couple*) The k-one-way couple pattern can be expressed using the temporal graph formalism with the following basic pattern and incremental rule:

$$\text{basic pattern} : KOWC(2) = (V_{KOWC_2}, E_{KOWC_2})$$
$$V_{KOWC_2} = \{u, v\}$$
$$E_{KOWC_2} = \{(v, u, t_i) | i \in \mathbb{N} \wedge i \leq 2 \wedge t_i < t_{i+1}\}$$

$$\tag{5}$$

$$\text{incremental rule} : KOWC(k + 1) = (V_{KOWC_{k+1}}, E_{KOWC_{k+1}})$$
$$V_{KOWC_{k+1}} = V_{KOWC_k}$$
$$E_{KOWC_{k+1}} = E_{KOWC_k} \cup \{(v, u, t_{k+1}) | t_k < t_{k+1}\}$$

In the k-one-way couple basic pattern, we find only two nodes u and v and two edges connecting them. However, differently from the ping-pong, both edges have the same source, that is node v, and the same destination, node u, as shown in Fig. 6. The incremental rule, which adds another edge from node v to node u to the pattern, can be used to understand at which extent the communication happens unilaterally.

This pattern, combined with the k-ping-pong, is used in our scenario to model a stalker/stalked behaviour. Indeed, all the edges model interactions happening in the same direction: from the stalker to the stalked. Differently from the k-ping-pong motif, there is no interaction going back from the stalked to the stalker, and differently from the k-in-star, the source of the interactions are not different nodes but it's the very same one (Fig. 6b).

5 The dataset

In this work, we study five specific communication patterns in Online Social Groups by evaluating a set of Facebook groups composing of 17 different groups divided into 5 categories [8,18,19,33]. All the groups have been chosen at random into the list of Facebook groups. We use a crawler due to the limitation of Facebook API to retrieve information about groups.

Fig. 6 The k-one-way couple pattern. **a** shows the basic pattern, while **b** shows how the incremental rule increases the size of the pattern
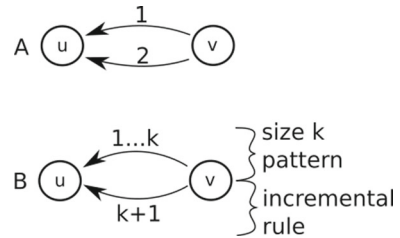
We registered an account in Facebook, and we joined these 17 closed groups after the acceptance by the administration of each group. The HTTP-crawler, which relies upon Selenium[1] to automate browser actions, periodically retrieve the following set of information from a specified Facebook group:

– *Members* We retrieved the list of the users participating in the group.
– *Interactions* We collected interactions occurred between the members of the groups. All interactions collected include posts, comments, and replies. Posts, comments and replies have a timestamp associated.

In particular, our crawler application was able to collect interactions of about 381,339 members belonging the 17 Facebook groups. We classify each group in one of five categories, depending on the description of the groups:

– *Education* It consists of Facebook groups discussing topics related to school or university.
– *Sport* It includes groups of users interested in popular sport activities such as football, tennis, or gym
– *Work:* It contains groups of users focused on business, job search, and companies.
– *Entertainment* It includes groups focused on media contents, such as film, music, or musicians.
– *News* This category contains groups of users interested in news, debates, and political discussions.

In order to help the readers in assessing the characteristics and the nature of the collected groups, Table 1 contains the real names, as well as the categories, of the groups which have been investigated in this study.

5.1 Data statistics

Table 2 summarizes the main characteristics of each group by showing: the number of consecutive days on which our crawler collected the interactions (*Days*), the number of group's member (*Members*), the date of both the first (*min Date*) and the last post (*max Date*) retrieved by our crawler, the total number of posts retrieved from the group in the monitored period (*Posts*), the number of members who are author of at least a post (*Authors*), and the average number of posts published each day by group members (*Post / day*).

The collected data indicate that our crawler was able to collect the activities performed by members over a period lasting longer than 365 days for Ed1, Ed3, S2, N3, W1, W2, and W4. Instead, other groups, in particular S3 have a high daily activity, and due to the overloading of the crawler (memory consumption), we were able to collect the activity of 28 days. The number of members of the groups are very heterogeneous, the maximum being

---

Table 1 General description of the selected Facebook's groups

GroupId	Group Name	Category
Ed1	Programmatori Need For Nerd - Code Community	Education
Ed2	Studenti in Crisi	Education
Ed3	AIRI (associazione internazionale ricercatori italiani)	Education
S1	Psiche e Benessere	Sport
S2	Pub del Nuoto	Sport
S3	Palestra Gruppo Ufficiale	Sport
W1	Italian Startup Scene	Work
W2	Sistemisti e Amministratori di Reti - Italiani	Work
W3	Malta Temp work	Work
W4	Discussion Cafe Jobs	Work
En1	Thriller storici e dintorni	Entertainment
En2	Horror House	Entertainment
En3	Scimmie Cinefile Official	Entertainment
En4	Chitarra Acustica	Entertainment
N1	Gossip, Anticipazioni, News: il meglio della TV	News
N2	Discutiamo di politica e attualitá	News
N3	Quick Books Tips and Tricks	News

Table 2 General description of the Facebook groups

| GroupId	Days	Members	min Date	max Date	Posts	Authors	Posts / day
Ed1	388	10,643	01/01/17	24/01/18	3555	1180	9.162
Ed2	317	46,016	06/04/17	18/02/18	5271	2643	16.628
Ed3	393	21,195	25/01/17	22/02/18	5060	1499	12.875
S1	249	35,671	27/08/17	03/05/18	5588	1562	22.441
S2	370	3,589	04/02/17	09/02/18	708	331	1.914
S3	28	107,459	13/02/18	14/03/18	6353	2754	226.893
W1	406	26,901	02/01/17	12/02/18	1444	800	3.557
W2	418	4,925	04/01/17	26/02/18	945	337	2.261
W3	318	25,257	13/06/17	27/04/18	4809	1140	15.122
W4	485	12,151	03/01/17	04/05/18	2651	374	5.466
En1	130	6,941	30/09/17	08/02/18	5009	915	38.531
En2	123	3,458	22/10/17	23/02/18	3777	319	30.707
En3	120	2,324	02/01/18	03/05/18	4904	1836	40.867
En4	178	9,392	09/09/17	06/03/18	3543	1041	19.905
N1	111	49,762	07/10/17	28/01/18	156	2	1.396
N2	91	11,663	08/11/17	07/02/18	3466	543	37.330
N3	406	3,992	02/01/17	12/02/18	1133	565	2.791
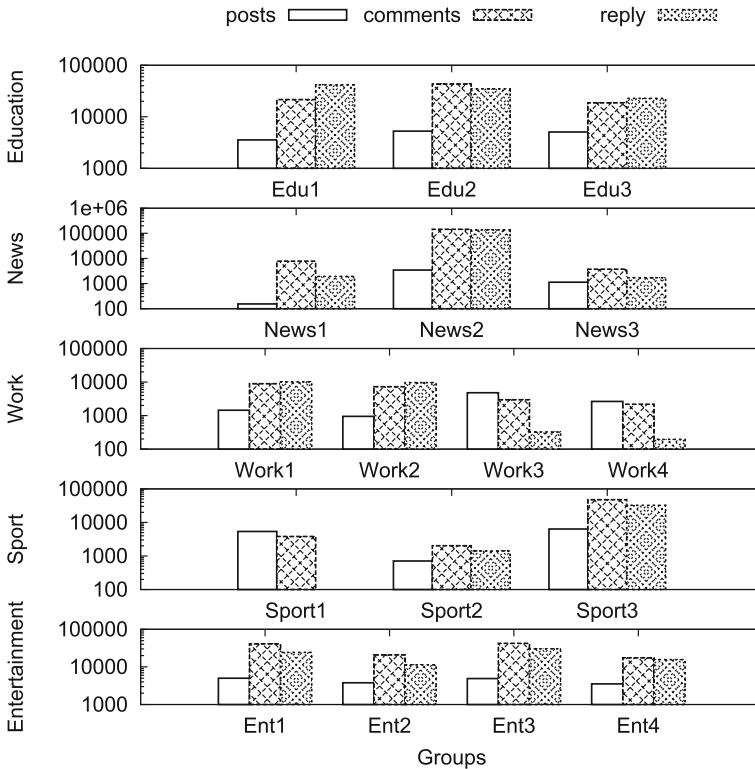
Fig. 7 Number and types of the interactions

107,459 users for S3, while the group En3 has the fewest number of members with only 2,324. The groups expose a high activity level, in fact the number of posts collected during the monitored period is higher than 4,000 for the majority of the groups. For instance, the group S3 has more than 6,000 published posts despite the fact that the monitored period lasts 28 days only. In general, a very low fraction of the group members are also author of at least one post collected during the monitored period (at most 10% of the group members). However, the collected groups expose higher variation in the number of distinct authors, and this value is positively correlated with the total number of posts (Pearson Correlation R = 0.64). The average number of posts per day computed by considering all the groups is equals to 8. However, its value depends on the social activity of the members, and it ranges between 1.914 posts (for group S2) and 226 posts (for group S3), while the groups of the Work category expose the least activity level.

We investigate in more detail the amount and the types of interactions collected from each group by showing in Fig. 7 the total number of posts, comments, and replies performed by members of the different groups. As shown by the figure, the groups expose different characteristics. The groups of the Education category have very similar number of posts but they expose different numbers of comments and replies. In general, the members of groups Ed1 and Ed3 interact mainly by using replies while group Ed2 exhibits a large portion of comment interactions. We notice that also groups of the Sport category expose similar characteristics where the majority of the interactions are comments. Instead, the groups of

the Work category exhibit two different trends: the groups W1 and W2 consist of few posts with a high number of comments and replies while the group W3 and W4 consist of several posts having a small number of comments and replies. Finally, the groups of the News and Entertainment category expose a similar communication pattern where members mainly interact by using comments, replies, and posts. As final remark, we can notice that, in most cases, the number of comments/replies is larger than the number of posts and, in many cases, the number of comments/replies is comparable.

6 Experimental results

In this section, we present experimental results about the detection of the Incremental Communication Patterns defined in sect. 4 in the scenario defined in Sect. 3. Experiments were carried out by considering the interactions among users belonging to the 17 Facebook groups presented in Sect. 5. The interactions were sorted based on their time label that represents the time when the interaction was received by the Facebook servers. Having the timestamp of each interaction is useful to capture the evolution of interactions of the groups. The sorted interactions are then used as input for a naive algorithm which detects the five Incremental Communication Pattern, i.e. the k-chain, the k-in/out-star, the k-ping-pong, and the k-one-way couple. For the sake of clarity, our goal is not to propose an efficient algorithm for the discovery of the Incremental Communication Patterns, rather we are interested in the study of the patterns, in the scenario of OSGs, where the communication between people is not yet studied in depth.

An essential property of our approach is that we are not searching for fixed patterns, but we are searching for the biggest pattern that appears. This is crucial to understand the property of a communication pattern in OSGs. In practice, our approach works in this way: We start from the first interaction of the stream and we try to build a basic pattern. Once a basic pattern is found, more edges are added to the instance of the pattern by applying recursively the incremental rule. The whole process stops when the next interactions fall out of the given time window, and the largest pattern is returned, or no pattern is returned in the unfortunate case in which not even the basic pattern was detected. Once the detection process is finished for the patterns starting with the first interaction, we iterate the process on the second interaction, and so on, until no more interactions can be found. For what concerns this study, we decided to filter out the instances of the patterns that match the basic pattern (except for the k-ping-pong). This decision was driven by the fact that in the particular scenario in which we apply the concept of Incremental Communication Pattern in this work, we understand that very short instances of some patterns appear too much frequently to be considered as relevant communication patterns.

As we saw in Sect. 5, the activity of the 17 groups is not homogeneous, and because of this it was not so easy to define a single length for a time window that was significant for all the groups. Indeed, a good time window length for an average group may be too short for groups with low activity and too long for groups with a lot of activity. Therefore, we decided to identify different time windows for each group studying the distribution of the interaction inter-arrival time (i.e. the amount of time passing between two successive interactions in all the group activity). Table 3 shows the results obtained by evaluating the inter-arrival time. In the group S3, which is the group in which the interaction activity is frequent, the average inter-arrival time is 31s while in W4 is 2251s. This justifies why we are not able to fix a specific time window for all groups.

Table 3 Statistical measures and quartiles for each group of the interaction inter-arrival time

Group	Min	Max	Avg	First qtl	Second qtl	Third qtl
Ed1	0	246,490	528	28	77	243
Ed2	0	194,612	337	24	68	198
Ed3	0	162,938	818	58	172	582
S1	0	67,336	132	17	45	109
S2	0	618,038	9,340	113	653	3,627
S3	0	19,336	31	5	14	30
W1	0	1,062,459	1,843	75	265	1,045
W2	0	264,094	2,142	65	213	993
W3	0	159,599	8,307	316	2,426	8,470
W4	0	776,642	17,550	265	2,251	13,601
En1	0	25,579	173	26	63	152
En2	0	28,867	331	35	95	266
En3	0	5,002,662	143	12	30	71
En4	0	40,659	465	47	129	375
N1	0	314,481	998	18	94	438
N2	0	120,669	28	4	11	26
N3	0	325,624	6,407	89	338	3,187

Since there was no way to define a single time window for all groups a priori, we decided to study more in detail the interactions inter-arrival times. Figure 8 shows the CDF (Cumulative Distribution Function) of the interaction inter-arrival times for all the groups in the dataset, divided by category: 8a for the education groups, 8b for the sport groups, 8c for the work groups, 8e for the news groups, and 8d for the entertainment groups. Interestingly enough, all the CDFs have a similar shape, with the only difference of being shifted compared to each other. S2, W3, W4, and N3 are the only groups with a sensibly different, smoother shape, which causes the inter-arrival times to be spread over a much wider range of values. These differences are caused by a general inactivity of the users in the groups, as confirmed by Table 3 where we can see that these 4 groups have the highest average and median interaction inter-arrival time. In particular, the groups of the Education and Sport categories expose a very low median inter-arrival time among interactions while the groups of the Work category have the largest median inter-arrival time. To be able to take into account this wide range of activity in the groups, we decided to fix different time window for each group according to the interaction inter-arrival times. In order to catch interaction patterns happening in short, medium and long time, we decided to fix three different time windows for each group. The three time windows are set equal to the first three quartiles of the interaction inter-arrival time distribution for each group. Table 3 shows the numerical values chosen for each time window. Again, most groups show a very similar trend, in fact the ratio between a second and first quartile and the ratio between third and second quartile is almost always between 2.5 and 3.5. The only groups not showing this characteristic are S2, W3, W4, and N3 which have been observed to be the ones with the lowest activity.
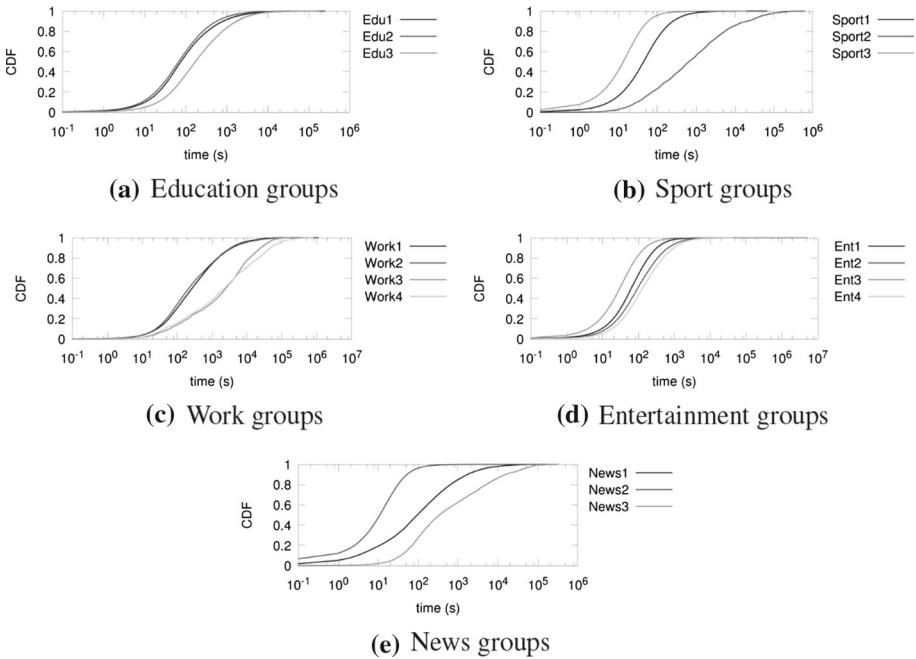
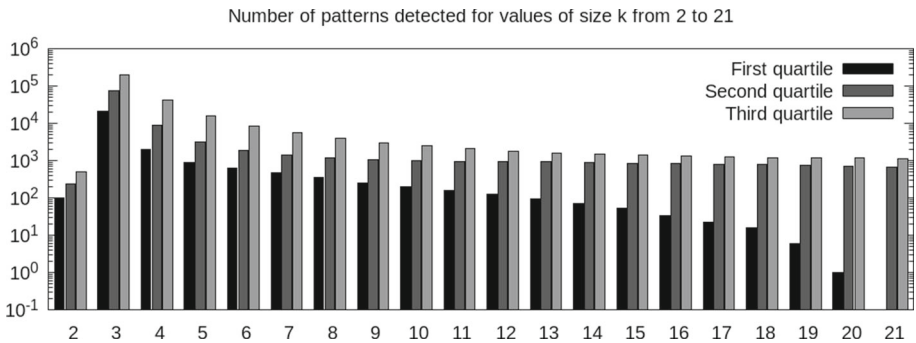**Fig. 8** CDF of the time passed between subsequent interactions

Fig. 9 The number of Incremental Communication Patterns for each value of k, up to 21

6.1 The recurring largest size evaluation

The first analysis performed concerns the study of the most frequent Incremental Communication Pattern maximal size which will let us understand what is the typical communication pattern length, that is the k value of the pattern returned. What we expect to find is the vast majority of patterns for low values of k (from 2 to 5), but we also expect to find larger patterns ($k > 5$).

Figure 9 depicts the number of Incremental Communication Patterns detected, for size of the pattern k up to 21, by using as time window length the first, the second, and the third quartile of the distribution of interaction inter-arrival times, summarized in Table 3. In general, we can observe that increasing the size of the time window has a significant difference on
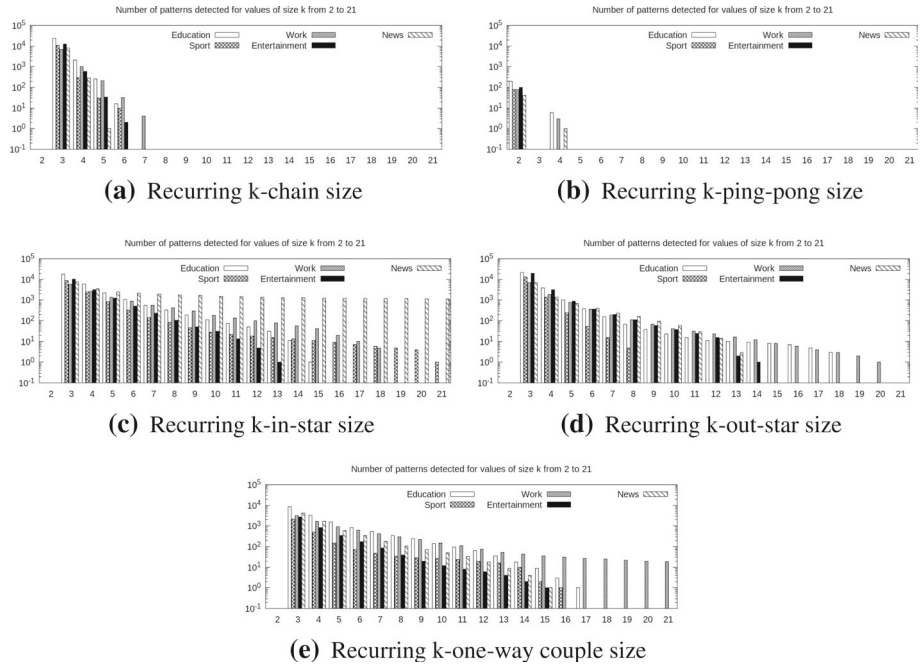
(a) Recurring k-chain size

(b) Recurring k-ping-pong size

(c) Recurring k-in-star size

(d) Recurring k-out-star size

(e) Recurring k-one-way couple size

Fig. 10 Most common size of the Incremental Communication Patterns studied in the groups

the total number of patterns identified in the groups. In particular, fixed the value k of the pattern, a larger number of patterns are identified as the size of the time window increases, as expected. This intuitively happens because the interactions can occur throughout a larger time span and still be included in the same pattern. Moreover, as expected, the most frequent maximal length of the patterns is 3, regardless of the time window, which tells us that the most frequent patterns of communication happens among a very small number of nodes. However, the plot also shows that there is a relevant number of patterns with non-standard length which are worth to be investigated, also in the shortest time window (the darkest one in the plot). In fact we see that, although much less common, we are able to detect patterns with 10 and more interactions, confirming our idea that one cannot restrict the search to standard and small patterns.

Figure 10 shows the distribution of the number of biggest Incremental Communication Patterns in the different group categories with k ranging from 2 to 21, one plot per pattern. Results are very interesting and are partially meeting our expectations. In general, we see that the most common values of k is 3 in all categories for all pattern, except for the k-ping-pong for which the most common k is 2, directly followed by 4 (we recall that the k-ping-pong pattern is defined only for even k). Nevertheless, we notice that the most common value of the size k always corresponds to the shortest pattern instances we consider: 2 for the k-ping-pong and 3 for all the other patterns.

Examining more in detail the plots one by one, we observe different behaviours for different patterns. Indeed, for two patterns, namely the k-chain (Fig. 10a) and the k-ping-pong (Fig. 10b), we see that the largest k observed is 7 for the former and 4 for the latter. The situation is quite different if we consider the other three patterns: k-in-star (Fig. 10c)), k-out-star (Fig. 10d), and k-one-way couple (Fig. 10e). Considering these patterns, we see that in almost

all categories we observe tens of pattern instances with size equal to 10, and even more interesting is the fact that in some cases the longest patterns exceed 20 in length.

Finally, we also observe that each category show a peculiar number of pattern for each different value of k. For example, concerning the number of k-in-stars (Fig. 10c), we see that the largest pattern for groups in the Entertainment and Education categories have $k = 13$ and $k = 15$, respectively, as largest sizes, while for the Sport category we have $k = 21$ and for the News category we count more than 1000 patterns with $k = 21$. A rather different situation can be seen for the k-out-star pattern (Fig. 10d). In this situation, we observe that the largest patterns in the Sport and News categories have $k = 8$ and $k = 13$, respectively, while Education and Work categories reach $k = 18$ and $k = 20$ respectively.

Overall, this first analysis suggests us that some communication patterns do not develop much in length, while others show a much more complex structure that is worth deeper investigation. We also observed that different categories show different number of patterns, therefore to have a more accurate view, we also plan to make further analyses at group granularity, rather than category granularity.

6.2 The absolute largest size evaluation

The main reason we introduced the Incremental Communication Patterns was to overcome the limit given to the number of actors and the number of interactions happening between them, which is typical of the motifs. To show the importance of removing this limitation, we decided to dedicate our attention to the study of the absolute maximal size of the patterns reached in each group separately.

Figure 11 shows, for each pattern, the maximum value of size k that has been discovered on all the groups by using different temporal window lengths: namely the first (11a), the second (11b), and the third (11c) quartile of the interaction inter-arrival distributions (Fig. 11d). As we can see from the Figures, the size of the time window highly affects the absolute maximal size of the patterns. Indeed, considering the results obtained with the shortest temporal window in Fig. 11e we see that the most common largest patterns have $k = 4$ and $k = 5$. We also notice few outliers: the largest k-out-star in group W1 with $k = 8$, the largest k-out-star in group W3 with $k = 10$, and the largest k-in-star in group N1 with $k = 20$. It is also worthwhile to notice that the k-ping-pong is a very specific pattern, and some groups (W3, W4, and N1) do not show any of them using the shortest window length.

Moving to the second time window length (results in Fig. 11b), we see an overall increase in the size of the largest patterns, as expected. The increase is much more highlighted for the k-one-way couple, and many groups have the largest k-one-way couple size doubled with respect to the previous window length. The absolute largest pattern is still a k-in-star in the group N1, whose size is tripled with respect to the shortest time window length, reaching $k = 62$; the second largest is a k-one-way couple in the group W4 with $k = 16$.

If we analyse the results for the third and longest time window (Fig. 11c), we see again an increase of the size of the largest patterns, especially for k-one-way couple and k-in-star. Concerning the k-one-way couple pattern, we observe two opposite behaviours: in some groups the maximal size is almost the same as in the previous time window (S2, S3, W3, W4, En1, and En2), while in other cases the size increases sensibly (Ed1, S1, W1, W2, and En3). Considering the k-in-star we observe that the increase of the largest size ranges from almost twice to three times with respect to the previous time window length in all groups, with only few exceptions: Ed1, En3, En4, N2. The absolute largest pattern is still a k-in-star in the N1 group, reaching the stunning size of $k = 170$; other large patterns are the k-one-way couple
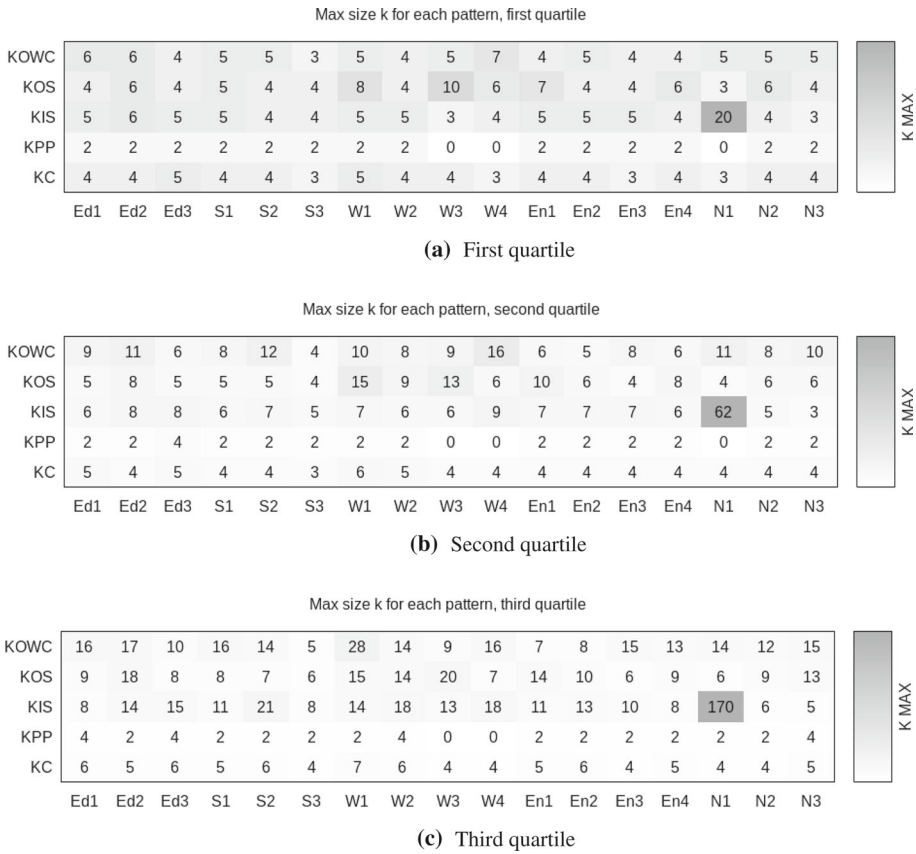
Max size k for each pattern, first quartile

	Ed1	Ed2	Ed3	S1	S2	S3	W1	W2	W3	W4	En1	En2	En3	En4	N1	N2	N3
KOWC	6	6	4	5	5	3	5	4	5	7	4	5	4	4	5	5	5
KOS	4	6	4	5	4	4	8	4	10	6	7	4	4	6	3	6	4
KIS	5	6	5	5	4	4	5	5	3	4	5	5	5	4	20	4	3
KPP	2	2	2	2	2	2	2	2	0	0	2	2	2	2	0	2	2
KC	4	4	5	4	4	3	5	4	4	3	4	4	3	4	3	4	4

(a) First quartile

Max size k for each pattern, second quartile

	Ed1	Ed2	Ed3	S1	S2	S3	W1	W2	W3	W4	En1	En2	En3	En4	N1	N2	N3
KOWC	9	11	6	8	12	4	10	8	9	16	6	5	8	6	11	8	10
KOS	5	8	5	5	5	4	15	9	13	6	10	6	4	8	4	6	6
KIS	6	8	8	6	7	5	7	6	6	9	7	7	7	6	62	5	3
KPP	2	2	4	2	2	2	2	2	0	0	2	2	2	2	0	2	2
KC	5	4	5	4	4	3	6	5	4	4	4	4	4	4	4	4	4

(b) Second quartile

Max size k for each pattern, third quartile

	Ed1	Ed2	Ed3	S1	S2	S3	W1	W2	W3	W4	En1	En2	En3	En4	N1	N2	N3
KOWC	16	17	10	16	14	5	28	14	9	16	7	8	15	13	14	12	15
KOS	9	18	8	8	7	6	15	14	20	7	14	10	6	9	6	9	13
KIS	8	14	15	11	21	8	14	18	13	18	11	13	10	8	170	6	5
KPP	4	2	4	2	2	2	2	4	0	0	2	2	2	2	2	2	4
KC	6	5	6	5	6	4	7	6	4	4	5	6	4	5	4	4	5

(c) Third quartile

Fig. 11 Largest k for each motif in each group varying the length of the time window

in group W1, k-in-star in S2, and k-out-star in W3. An interesting result, forewarned in Sect. 6.1, is that in two groups in the Work category, namely W3 and W4, we observe that there is no k-ping-pong at all, and in all other groups we observe only small values of k.

6.3 Groups correlation

Lastly, we decided to investigate the dependencies existing among the Incremental Communication Patterns of different groups by exploiting the Pearson correlation coefficient. We used a correlation matrix to show the correlation results where rows and columns represent the groups while the color of each matrix's cell denote the correlation among the corresponding groups where white, light grey and dark grey indicates negative correlation, no correlation, and positive correlation, respectively.

Figure 12 shows the correlation matrix between groups by considering the absolute pattern size k for all the patterns and for all the quartiles. The matrix does not clearly show very high correlation values, meaning that the groups show an heterogeneous environment of maximal length patterns. We can, however, roughly identify two clusters of groups. The first one is made of the following groups: Ed1, S1, En3, N2; the second one is made of the following
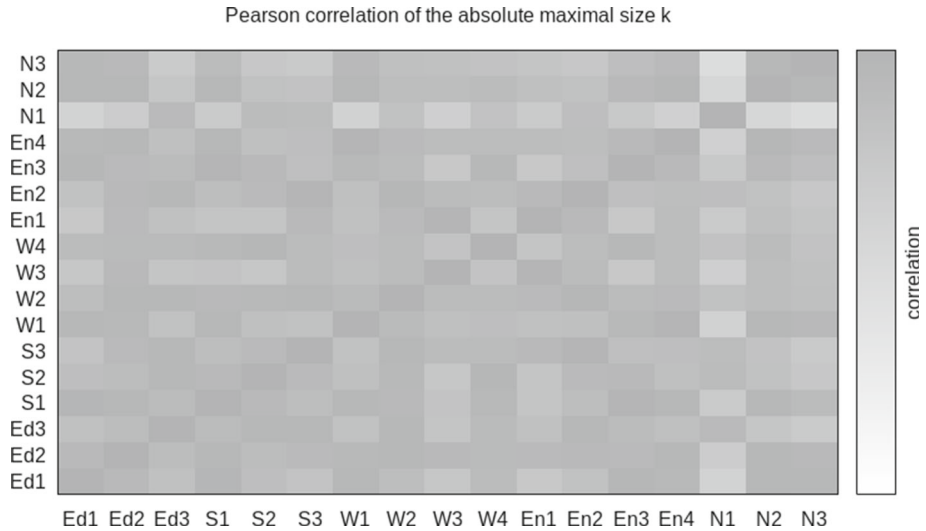
Pearson correlation of the absolute maximal size k

Fig. 12 Correlation matrix of the groups based on the absolute largest size of all the Incremental Communication Patterns and for all the quartiles

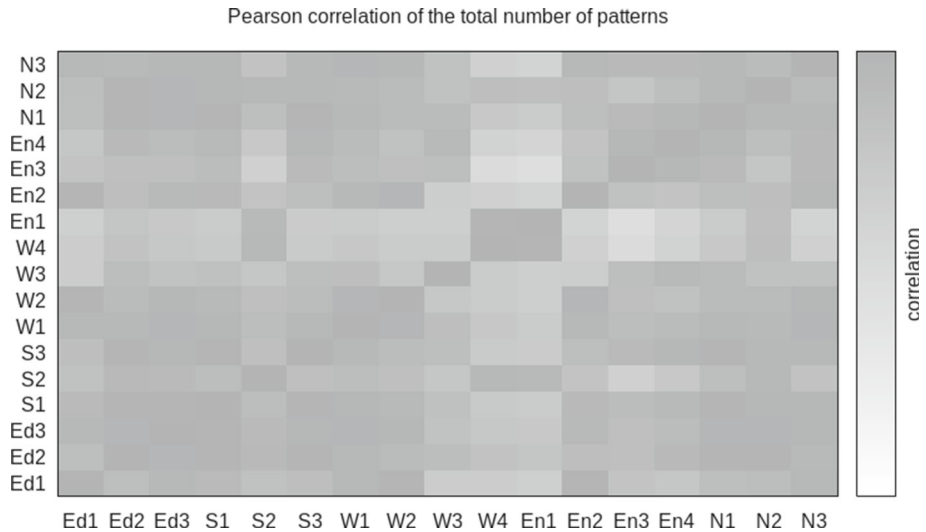Pearson correlation of the total number of patterns

Fig. 13 Correlation matrix of the groups based on the number of Incremental Communication Patterns found in each group

groups: Ed3, S3, W2, En2. The group N1 seems to have very low correlation values with other groups, showing a rather unique combination of maximal lengths of the patterns.

Seeing that there is no clear correlation in the maximal size of the Incremental Communication Patterns, we decided to study if there is correlation in the pattern count among the different groups (Fig. 13). Interestingly enough, the matrix clearly indicates the presence of two clusters of positive correlation. The first one among a large set of groups which consists of the groups of the Education and News categories, and the groups S1, S3, W1, W2, and

Pearson correlation of the number of patterns for values of size k available to all groups
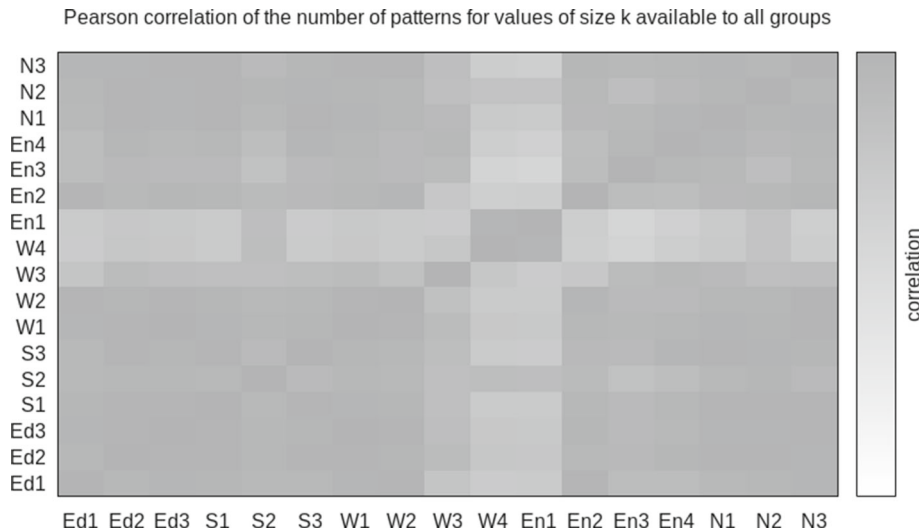
Fig. 14 Correlation matrix among groups which is based on the aggregated number of patterns, for all values of k, for all sizes of the temporal windows, and for all the types of patterns

En1. In the second cluster, we find the groups S2, W4, and En1. It is worth to notice that, while in the previous case there was no clear distinction, here we can easily identify two distinct clusters, meaning that groups can be characterized by the number of Incremental Communication Pattern detected. A very similar result is obtained if we restrict the correlation to the number of patterns only to values of k which are available to all groups, as we can see in Fig. 14. The only big difference here is that the group S2 does not seem to have high correlation with groups W4 and En2.

Despite having found that some of the groups show a similar number of patterns, or patterns of similar sizes, the correlations show clusters of groups with heterogeneous characteristics. In fact, we see that groups belonging to different categories and with different activity are clustered together and, alongside with that, groups with a similar activity belong to different clusters. This result confirms that the groups are highly heterogeneous, not only considering the number of users and their activity, but also considering the communication patterns that one can observe in them.

7 Conclusions and future works

In this paper, we defined and studied Incremental Communication Patterns to identify previously undetected communication structures in OSGs. In detail, we proposed and formalized the concept of Incremental Communication Pattern, together with the concept of size of the pattern, to study the communication structure appearing in fixed-length time windows. We proposed five Incremental Communication Patterns, which identify specific communication structures in OSGs, by the means of a basic pattern to identify the initial meaningful structure, and an incremental rule to be applied to the basic pattern to add more interactions to the pattern. We studied them by exploiting the interactions among users of a real Facebook dataset consisting of 17 Facebook Groups belonging to five different categories. The detection is also guided by a way of limiting the number of interactions that can be part of the pattern in

terms of the time passed from the first interaction of the pattern. Results show that, beyond simple patterns which involve only a fixed number of interactions and users, a relevant set of large patterns having a non-trivial number of components can be recognized. In particular, some real groups defined in Facebook expose very complex communication patterns which engage up to 170 members of the group. Out of the five Incremental Communication Patterns defined in this paper, we also see that only some of them are prone to show a complex structure which was, up to now, undetected due to the limitation of the motifs proposed in literature. As future works, we plan to propose other specific Incremental Communication Patterns and study them in the same context using more Facebook Groups, taking into account the list of popularity of them. Moreover, we plan to investigate a more accurate mathematical model which can be applied to slice the time more consistently, which takes into account the density of the interactions and which provides a variable-length time window.

Declarations

Andrea Michienzi got a Bachelor degree in December 2015 followed by a Master degree in October 2017 in Computer Science at the Università di Pisa. Starting from November 2017, he is a PhD student at the Department of Computer Science at the Università di Pisa, a position he is still holding at the present times. His main research interests are distributed systems and complex network analysis. More in detail, he is currently investigating techniques for community detection throughout time, focusing on exploiting these mechanisms for the development of Distributed Online Social Networks. Moreover, he is starting to investigate other blue-sky research topics, such as the dynamics of socio-economic networks and influencer and area-of-influence detection in online social networks.

Barbara Guidi is currently a postdoctoral researcher at the Department of Computer Science of the University of Pisa. She received her B.Sc. and M.Sc. in Computer Science from the University of Pisa, Italy, in 2007 and 2011, respectively. She received her Ph.D. degree in Computer Science from the University of Pisa, in 2015. She was a Co-Chair for the conference EAI GoodTechs 2017, and Co-Chair of several workshops. She has been involved in the TPC of several International conferences and workshops and has been a reviewer for relevant journals, such as IEEE Access, and Concurrency and Computation: Practice and Experience (CCPE). She received three Best Paper Awards: at the International Conference DCNET 2013, at the workshop LSDVE 2017, and at LSDVE 2018. Her current research interests include distributed systems, P2P networks, complex networks, Social Network Analysis, Decentralized Online Social Networks, dynamic community detection, and the Blockchain technology.

Laura Ricci received the M.Sc. in Computer Science from the University of Pisa in 1983 and the Ph.D. from the same University, in 1990. She is currently Associate Professor at the "Department of Computer Science","University of Pisa", Italy. Her research interests include distributed systems, peer-to-peer networks, cryptocurrencies and blockchains and social network analysis. In these fields, she has co-authored 100+ papers published on international journals and conference/workshop proceedings. She has served as programme committee member and chair of several conferences. In particular, she has been Program Chair of the 19th edition of "DAIS, International Conference on Distributed Applications and Interoperable Systems". She is the organizer of the LSDVE, Large Scale Distributed Environments workshop, held in conjunction with EUROPAR conference. She has been involved in several research projects and she is currently the local coordinator of the H2020 European Project "Helios: A contect aware Distributed Networking Framework".

Andrea De Salve is a computer science researcher at the Institute of Applied Sciences and Intelligent Systems (ISASI) of the National Research Council (CNR), Italy. He received his Ph.D. in Computer Science from the University of Pisa and he worked at IIT-CNR and University of Palermo as a junior researcher. His main research interests include security and privacy in distributed systems, complex network analysis, web mining, trust in the context of Decentralized System, Distributed Ledger Technology, P2P system, and Online Social Networks. In particular, he investigates protocols and algorithms for large scale distributed system, properties of the users' behaviour in Online Social Networks, and P2P solutions for decentralizing Online Social Networks.

References

1. Arndt J (1967) Role of product-related conversations in the diffusion of a new product. J Mark Res 4(3):291–295

2. Backstrom L, Kumar R, Marlow C, Novak J, Tomkins A (2008) Preferential behavior in online groups. In: Proceedings of the 2008 International Conference on Web Search and Data Mining, WSDM '08, pp 117–128

3. Bhattacharya A, Desai H, DeMarse TB, Wheeler BC, Brewer GJ (2016) Repeating spatial-temporal motifs of ca3 activity dependent on engineered inputs from dentate gyrus neurons in live hippocampal networks. Frontiers in neural circuits 10:45

4. Bickart B, Schindler RM (2001) Internet forums as influential sources of consumer information. J Interact Market 15(3):31–40

5. Braines D, Felmlee D, Towsley D, Tu K, Whitaker RM, Turner LD (2018) The role of motifs in understanding behavior in social and engineered networks. In: Next-Generation Analyst VI, vol. 10653, International Society for Optics and Photonics, p. 106530W

6. Casteigts A, Flocchini P, Quattrociocchi W, Santoro N (2012) Time-varying graphs and dynamic networks. Int J Parallel Emergent Distrib Syst 27(5):387–408

7. Creusefond J, Cazabet R (2017) Characterising inter and intra-community interactions in link streams using temporal motifs. In: Workshop on Complex Networks CompleNet, Springer, pp 81–92

8. De Salve A, Guidi B, Michienzi A (2018a) Studying micro-communities in Facebook communities. In: Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good, pp 165–170

9. De Salve A, Guidi B, Ricci L, Mori P (2018b) Discovering homophily in online social networks. Mobile Netw Appl 23(6):1715–1726

10. De Salve A, Mori P, Guidi B, Ricci L (2019) An analysis of the internal organization of facebook groups. IEEE Trans Comput Soc Syst 6(6):1245–1256

11. Echtermeyer C, Han CE, Rotarska-Jagiela A, Mohr H, Uhlhaas PJ, Kaiser M (2011) Integrating temporal and spatial scales: human structural network motifs across age and region of interest size. Front Neuroinformatics 5:10

12. Faraj S, Johnson SL (2011) Network exchange patterns in online communities. Organ Sci 22(6):1464–1480

13. Garg R, Smith MD, Telang R (2011) Discovery of music through peers in an online community. In: 2011 44th Hawaii International Conference on System Sciences, IEEE, pp 1–10

14. Garton L, Haythornthwaite C, Wellman B (1997) Studying online social networks. J Comput Mediat Commun 3(1):JCMC313

15. Gross R, Acquisti A (2005) Information revelation and privacy in online social networks. In: Proceedings of the 2005 ACM workshop on Privacy in the electronic society, ACM, pp. 71–80

16. Guidi B, Michienzi A, Rossetti G (2017) Dynamic community analysis in decentralized online social networks. In: European conference on parallel processing, Springer, pp 517–528

17. Guidi B, Michienzi A, Rossetti G (2019) Towards the dynamic community discovery in decentralized online social networks. J Grid Comput. 17(1):23–44

18. Guidi B, Michienzi A, De Salve A (2020) Community evaluation in Facebook groups. Multim Tools Appl 79(45–46):33603–33622

19. Guidi B, Michienzi A, Ricci L, Ambriola V (2021) Analysing Dunbar Circles in Facebook Groups. In: 2021 IEEE 18th Annual Consumer Communications Networking Conference (CCNC), pp 1–6. https://doi.org/10.1109/CCNC49032.2021.9369495

20. Hidalgo CA, Rodríguez-Sickert C (2008) The dynamics of a mobile phone network. Phys A Stat Mech Appl 387(12):3017–3024

21. Holland PW, Leinhardt S (1974) The statistical analysis of local structure in social networks

22. Holme P, Saramäki J (2012) Temporal networks. Phys Rep 519(3):97–125

23. Hulovatyy Y, Chen H, Milenković T (2015) Exploring the structure and function of temporal networks with dynamic graphlets. Bioinformatics 31(12):i171–i180

24. Jackson A, Yates J, Orlikowski W (2007) Corporate blogging: Building community through persistent digital talk. In: 2007 40th annual hawaii international conference on system sciences (HICSS'07), IEEE, pp 80

25. Jiang YF, Li CP, Han JZ (2009) Stock temporal prediction based on time series motifs. In: 2009 International conference on machine learning and cybernetics, vol 6, IEEE, pp 3550–3555

26. Jurgens D, Lu TC (2012) Temporal motifs reveal the dynamics of editor interactions in wikipedia. In: Sixth International AAAI Conference on Weblogs and Social Media

27. Karnik M, Oakley I, Venkatanathan J, Spiliotopoulos T, Nisi V (2013) Uses & gratifications of a facebook media sharing group. In: Proceedings of the 2013 conference on Computer supported cooperative work, ACM, pp 821–826

28. Kovanen L, Karsai M, Kaski K, Kertész J, Saramäki J (2011) Temporal motifs in time-dependent networks. J Stat Mech Theory Exp 2011(11):P11005

29. Kovanen L, Kaski K, Kertész J, Saramäki J (2013) Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences. Proce Natl Acad Sci 110(45):18070–18075

30. Liu K, Cheung WK, Liu J (2013) Detecting stochastic temporal network motifs for human communication patterns analysis. In: Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining, ASONAM '13, ACM, New York, NY, USA, pp. 533–540. https://doi.org/10.1145/2492517.2492525

31. McGovern A, Rosendahl DH, Brown RA, Droegemeier KK (2011) Identifying predictive multi-dimensional time series motifs: an application to severe weather prediction. Data Min Knowl Discov 22(1–2):232–258

32. Mellor A (2018) The temporal event graph. J Complex Netw 6(4):639–659

33. Nasti L, Michienzi A, Guidi B (2021) Discovering the Impact of Notifications on Social Network Addiction. In: From Data to Models and Back. Springer International Publishing, Cham, 72–86

34. Paranjape A, Benson AR, Leskovec J (2017) Motifs in temporal networks. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, ACM, pp 601–610

35. Pfitzner R, Scholtes I, Garas A, Tessone CJ, Schweitzer F (2013) Betweenness preference: Quantifying correlations in the topological dynamics of temporal networks. Phys Rev Lett 110(19):198701

36. Purohit H, Ruan Y, Fuhry D, Parthasarathy S, Sheth AP (2014) On understanding the divergence of online social group discussion. ICWSM 14:396–405

37. Redmond U, Harrigan M, Cunningham P (2012) Identifying time-respecting subgraphs in temporal networks. In: Proceedings of the european conference on machine learning and principles and practice of knowledge discovery in databases, pp 51–63

38. Saramäki J, Holme P (2015) Exploring temporal networks with greedy walks. Eur Phys J B 88(12):334

39. Schneider CM, Belik V, Couronné T, Smoreda Z, González MC (2013) Unravelling daily human mobility motifs. J R Soc Interface 10(84):20130246

40. Sproull L (2004) Online communities. In: The Internet Encyclopedia. American Cancer Society. https://doi.org/10.1002/047148296X.tie128

41. Tang J, Musolesi M, Mascolo C, Latora V (2010) Characterising temporal distance and reachability in mobile and online social networks. ACM SIGCOMM Comput Commun Rev 40(1):118–124

42. Tarbush B, Teytelboym A (2012) Homophily in online social networks. In: International Workshop on Internet and Network Economics, Springer, pp 512–518

43. Wellman B, Rainie L (2012) Networked. MIT Press, Cambridge

44. Wu J, Liu J, Chen W, Huang H, Zheng Z, Zhang Y (2020) Detecting mixing services via mining bitcoin transaction network with hybrid motifs. arXiv preprint arXiv:2001.05233

45. Xuan Q, Fang H, Fu C, Filkov V (2015) Temporal motifs reveal collaboration patterns in online task-oriented networks. Phys Rev E 91(5):052813

46. Zhang YQ, Li X, Xu J, Vasilakos AV (2015) Human interactive patterns in temporal networks. IEEE Trans Syst Man Cybern Syst 45(2):214–222

47. Zhao Q, Tian Y, He Q, Oliver N, Jin R, Lee WC (2010) Communication motifs: a tool to characterize social communications. In: Proceedings of the 19th ACM international conference on Information and knowledge management, ACM, pp 1645–1648

48. Zignani M, Quadri C, Del Vicario M, Gaito S, Rossi GP (2018) Temporal communication motifs in mobile cohesive groups. In: Complex Networks & Their Applications VI, Springer International Publishing, pp 490–501