



# Identification of Patterns in Blogosphere Considering Social Positions of Users and Reciprocity of Relations

Krzysztof Rudek<sup>(✉)</sup> and Jarosław Koźlak

Faculty of Computer Science, Electronics and Telecommunications,  
AGH University of Science and Technology, Kraków, Poland  
{rudek, kozlak}@agh.edu.pl

**Abstract.** The aim of the paper is to identify and categorize frequent patterns describing interactions between users in social networks. We consider a social network with already identified relationships between users which evolves in time. The social network is based on the Polish blog website pertaining on socio-political issues [salon24.pl](http://salon24.pl). It consists of bloggers and links between them, which result from the intensity and characteristic features of posting comments. In our research, we discover patterns based on frequent and fast interactions between pairs of users. The patterns are described by the characteristics of these interactions, such as their reciprocity, the relative difference between estimates of global influence in the pairs of users participating in the discussions and time of day of the conversation. In addition, we consider the roles of system users, determined by the number of interactions initiating discussions, their frequency and the number of strong interactions in which users are involved. We take into account how many such intense conversations individual users participate in.

**Keywords:** Social network analysis · Social relations  
Identification of social patterns · Blogosphere

## 1 Introduction

The aim of the work is to identify and analyze patterns frequently occurring in the social network evolving in time. Finding patterns allows to determine which relationships in the social network are particularly important and how to better understand their nature, which in turn may lead to a better understanding of the behavior of the entire network and how individual users affect each other and the entire community. Such information may help to understand ways of propagation ideas, which can be used to promote products or political views, and also to identify, for example, unlawful activities, their sources and ways of propagating them.

The analysis of blogosphere can be a useful area for testing solutions to such problems, due to open access to this data, a large scope of available data and the ability to observe the interactions between individual users in a long (multi-year) time horizon.

The graph evolving in time is described by a set of subsequent graphs represented its structure in subsequent time intervals. This allows us to analyze many graphs with a similar set of vertices and extract such an element in the graph (both from the point of

view of the existence of individual relationships, sets of relationships and attributes assigned to vertices and edges) which often occurs.

## 2 State of the Art

In this section we will discuss issues related to the states of a social networks (with data coming usually from networking portals or phone calls), characteristics of relations between users in such portals and types of frequent patterns which may be identified in such graph.

Different names are used for frequently present elements in networks: patterns [10], cascades [5], persistent cascades [12], graphlets [6], fundamental structures [9], motifs [8] or sub-graphs [11]. In this paper we will use the name “patterns”. Considered patterns are usually small bricks building networks containing several (usually 3–4) nodes.

The authors of [6] analyze patterns in both statics and dynamics networks. In [8] are analyzed temporal patterns with events which do not overlap in time and considered patterns allow to describe a sub-network topology and a sequence of events. Considered data sets contains information about phone calls. In [10] are studied patterns in data coming from blogosphere and they describe relations between posts or blogs. In [12] the authors analyze patterns containing few nodes in networks created from the data about phone calls. In [14] are considered casual paths of events, temporal distances between events are calculated as well as their distribution and correlations. Measures of closeness centrality are calculated for nodes in analyzed graphs. The dataset concerns phone calls, air transport networks and random graphs. In [7] are identified motifs and dynamic motifs of the network which appear in the network with highest frequency. The authors analyze dynamic motifs mostly having 3 nodes and use the measure of the statistical significance comparing results for *Yahoo Answers* and *Flickr* with the results obtained for random graphs.

In [5] the patterns are considered for datasets gathered from Twitter and containing posts related to crisis situations.

Another approach is an identification of heavy subgraphs [1], and their analysis for time evolving networks. They are the sub—graphs with highest scoring which means having the highest weights of the connections.

Fundamental structures [9] are identified for temporal networks describing frequent meetings in graphs representing meeting of the persons. The authors consider different kinds of networks concerning meeting in various special locations and time periods.

In [4] dynamic patterns are used for the identification of contributions of nodes and paths in the information flow.

In the presented works the patterns were identified in different ways. What distinguishes our approach is to include in the patterns not only links between users and their topologies, but also additional attributes describing users and character of the relationship (absolute and relative social significances, timely distribution of interactions, their intensity, duration, etc.).

### 3 Methodology

In order to build relations between users in a community portal, all interactions between users are usually considered. In the case of the blogosphere, both the relationship between the author of a given comment and the author of the post in the context of which the comment is written, or between the author of the comment and the author of the commentary, which is commented, may be considered [3].

In our work, we focus on specific types of connections, which we call *strong relationships*, as we consider comments written quickly after the post or comment they respond to. We consider two different time intervals: of 5 min and 15 min. Another particularly important feature describing relationships is the reciprocity - if both users comment on their posts and comments.

The aim of the work is to distinguish patterns describing such strong and frequent connections occurring on the social portal. The patterns include the following parts:

- describing topology of relations between users (RT),
- containing attributes describing relationships (RA),
- social and mutual relations of users (SP).

Therefore, the analyzed pattern can be written as the following n-k

$$P = (RT, RA, SP) \quad (1)$$

#### Topology of relations between users

The topology is associated with the edges in the network, which are included in the considered pattern. The simplest pattern consists of two users/nodes and one edge, and on those we focus in this work. However, more complex patterns covering more users may also be considered.

The next main issue is whether the relationship is mutual and to what extent, i.e. how many posts and comments of each party meets the appropriate commentary of another member of the pattern.

#### Features of relations between users (RA)

Relations in the patterns can also be characterized by additional, detailed attributes describing the temporal characteristics of the interactions taking place between the users in question. In particular, the duration of their interactions and their volume are considered. Another important feature describing the relationship is the intensity of interactions at particular periods of the day.

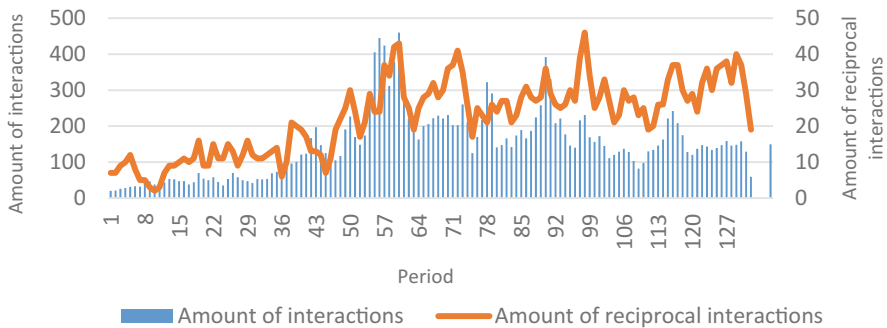
#### Social and mutual position of users (SP)

An important feature characterizing the relations between users are the social positions of the users participating in them, as well as the mutual relationship between these social positions. Different classes of relationships are the ones between users of high social importance, between users with insignificant social positions, and different between a user with a high social importance and a small social importance.

## 4 Description of Experiments

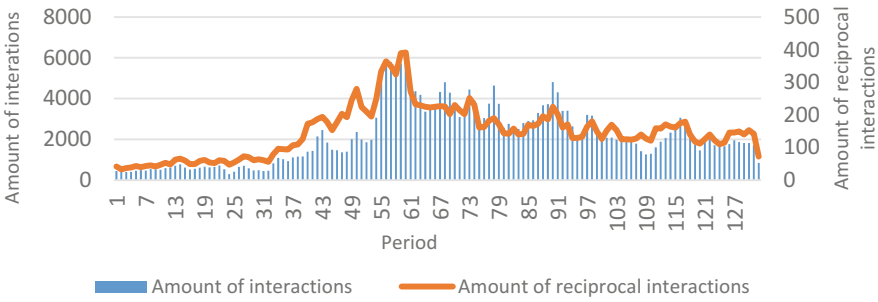
### 4.1 Characteristics of Interactions in Network

The social network that we used for our research is based on data from the [www.salon24.pl](http://www.salon24.pl) portal [3, 15]. Dataset was based on objects representing portal users, which are nodes in the network we built, as well as on objects representing posts and comments written by portal users. All objects are described with a set of attributes, which allowed to create an edge between the vertices of the graph and study how these edges change over time. The aim of our research is to find attributes of relations between portal users that will allow us to classify these relationships and find behavioral patterns. To better understand data and to find patterns based on interactions' attributes, we first need to look at amount of data and it's type. We want to know what is the quantitative dependence between the interactions in a given time interval for quick reactions in this range. Another feature we are considering at this stage is the same relation to fast interactions that were reciprocated equally quickly. The average number of interactions in the studied time intervals is 100000. The graphs below (Figs. 1 and 2) show how many fast interactions and mutual quick interactions have occurred, with the assumed delays of 5 min and 15 min.



**Fig. 1.** Amount of interactions and reciprocal interactions (5-min slot).

At the beginning of the graph analysis, it should be noted that the test data was collected from the moment when there was a large amount of interaction between agents in the social network, however it was still the initial stage of network development. Therefore, the increase in both the number of quick interactions between agents and mutual relationships is natural. Looking at Fig. 1. Presenting amount of interactions and reciprocal interactions in 5-min slot, three main peaks are noticeable - they all depend on important socio-political events. Bearing in mind the average number of interactions in the studied time intervals, we note that the number of fast interactions is low, between 20 and 460, with an average value of 149. On average, we report interrelations 22 over a given time interval, with a minimum value of 2 and a maximum of 46.



**Fig. 2.** Amount of interactions and reciprocal interactions (15-min slot).

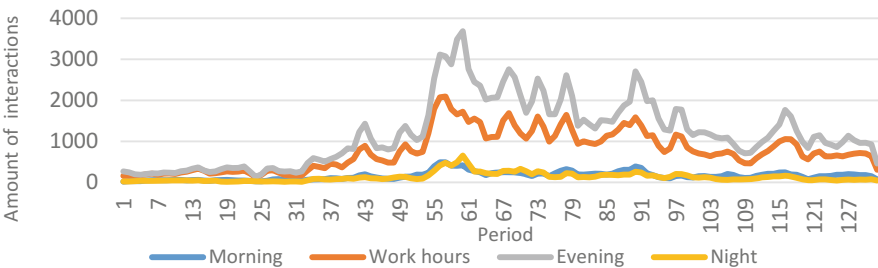
At a chart presenting 15-min time slot (Fig. 2) we again notice a gradual increase in both values. In contrast, only one summit is dominant, for which an important socio-political event is again responsible. The lack of very visible other peaks, which we can see in the case of the 5-min slot, can be explained by the lower overtones of these events in the country and in the world. We can easily see, however, that we are dealing here with many times greater representation of data. The number of interactions, classified as fast, increased to an average of 2151, with a minimum of 296 and a maximum of 5787. The number of interactions increased on average to 147, with a minimum of 33 and a maximum of 391.

Due to such a large discrepancy, we decided to base most experiments on a 15-min slot.

#### 4.2 Amount of Interactions During the Time of the Day

One of the important characteristics that we are considering when looking for patterns in a social network is the time of day when agents interact with each other. Based on the test set, we chose 4 times of the day, which we then used to classify the interaction. They are: morning (06.00 AM–10.00 AM), working hours (10.00 AM–06.00 PM), evening (06.00 PM–00.00 AM), night (00.00 AM–06.00 AM).

As can be seen in the graph (Fig. 3), the vast majority of interactions between agents takes place during business hours and in the evenings. The increases in the moments of the highest peaks are distributed proportionally. This proves a certain regularity in interactions, which can be used as an important factor to search for patterns in the social network.



**Fig. 3.** Amount of interactions during the time of day (15-min slot).

Based on the knowledge presented in the previous graphs (Figs. 2 and 3), it can be said that the distribution of the amount of rapid and mutual interactions between agents looks predictable (see Fig. 4). The increases are proportional and, as in the case of ordinary interactions, most often occur during working hours and in the evenings. It is worth mentioning that the distribution for this 5-min interval looks very similar. However for this characteristics as well as some in further research presented in next paragraphs, distribution of interactions for 5-min slot has not been presented, as, due to the small amount of interaction, in further research we focus on the 15-min interval. This decision was caused by the knowledge of the portal at once, where serious socio-political topics are raised, and often it takes more than 5 min to read the content published by the user.

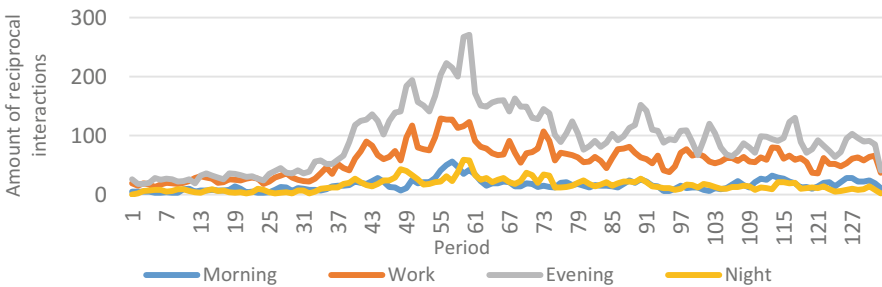


Fig. 4. Amount of reciprocal interactions during the time of day (15-min slot).

### 4.3 Cardinality of the Sets in Terms of the Number of Interacting Agents

Having knowledge about how fast and mutual interactions are and how the day the agents interact with each other, we want to check with how many agents they usually interact with. For this purpose, we have divided into groups of agents who interact with one, two, three etc. with other agents. Then, for ease of presentation, we made a grouping of results, which is shown in the graphs below.

In the case of a 5-min slot (Fig. 5), we received two groups. The first one includes those agents who interact with up to 5 other agents. This is the vast majority. The second group are agents who interact with more than 5 agents. It is a very small group, but from the point of view of looking for patterns, it is very important, allowing for the bolding of the characteristics of individual interactions.

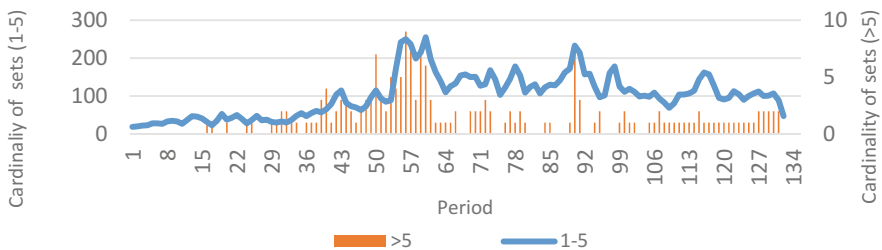
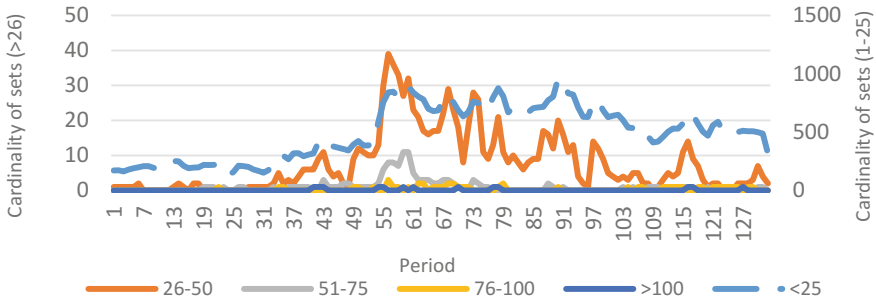


Fig. 5. Cardinality of the sets in terms of the number of interacting agents (5-min slot).



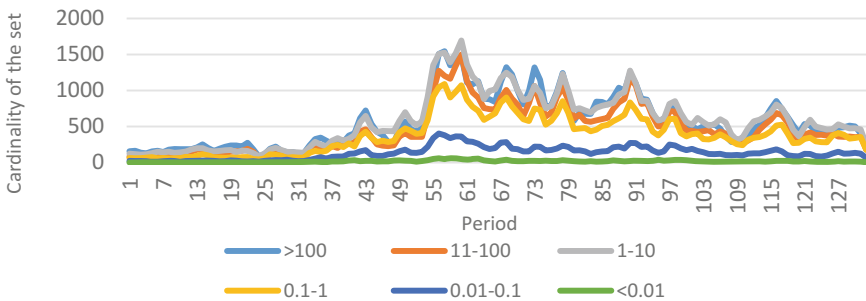
**Fig. 6.** Cardinality of the sets in terms of the number of interacting agents (15-min slot).

The characteristics for 15-min slot (Fig. 6) looks similar. The vast majority of agents interact with a group of several to 25 other agents.

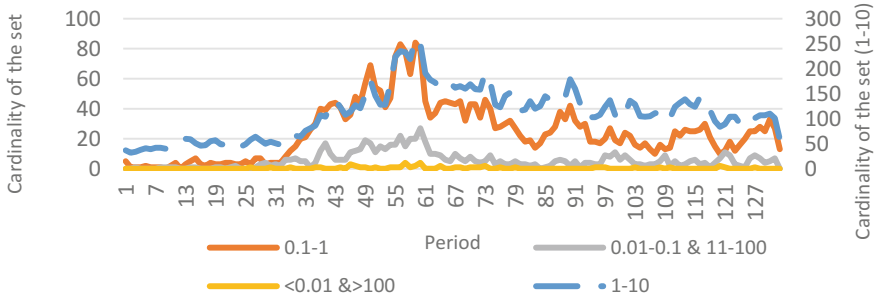
#### 4.4 Cardinality of the Sets in Terms of Agent’s In-Degree Ratio

Very important factor influencing the establishment of interaction between agents is their own characteristics. Agents are characterized by values such as in-degree and out-degree, describing how much the agent brings to the network and how the other members of the network are focused on his actions. We made an analysis that grouped together agent pairs, taking the ratio of the in-degree ratio of each pair as a factor. It turned out that we are dealing with six groups. Markings on the graph indicate the agent’s in-degree coefficient in relation to which the executed action takes place.

In the case of quick interactions, it turned out that the agents most often interact with those who have the in-degree parameter higher several times (see Fig. 7). The next in terms of population size was the one containing agents affecting over a hundred times more influential agents. It is worth noting that at the moments of the highest stitches, it is even the most numerous group. The groups of agents interacting with those whose in-degree parameter is higher several dozen times or a few times lower are not much smaller. On the other hand, agents rarely interact with those much less influential.



**Fig. 7.** Cardinality of the sets in terms of agent’s in-degree ratio (15-min slot).



**Fig. 8.** Cardinality of the sets in terms of agent’s in-degree ratio for reciprocal interactions (15-min slot).

The same division made for quick and mutual interactions has emerged four groups of agents (two overlap, because they are symmetrical) (see Fig. 8). It is very clearly visible that agents who interact quickly and interact with each other have very similar characteristics in terms of social network (in-line several times higher, or several times lower). Such situations are almost unnoticeable in the case of very large differences with the pair.

### 4.5 Studied Relationships

In order to check how the surveyed characteristics translate into the characteristics of the existing relationships, we examined the relationships between agents, characterized by the highest number of interactions. It turned out that in a small collection of 14 accounts, there were different types of relationships, both one-sided and mutual, lasting for a long time, as well as those with very frequent intervals, based on agents with similar characteristics as well as very diverse ones. The relations are described by the following characteristics: (A) in-degree ratio at the beginning of relationship, (B) in-degree ratio at the end of the relationship, (C) in-degree ratio amplitude, (D) – in-degree ratio noted when relationship was mutual, (E) Durability of unilateral relationship (periods when interaction was present), (F) Reciprocity [%], (G) – Reciprocity (periods when relationship was mutual), (H) Longest sequence of periods with interactions, (I) Longest sequence of periods without interactions, (J) Longest sequence of periods with mutual interactions, (K) Longest sequence of periods without mutual interactions, (L) Percentage of interactions taken in the morning [%], (M) Percentage of interactions taken during business hours [%], (N) Percentage of interactions taken in the evening [%], (O) Percentage of interactions taken in the night [%].

Looking at the results presented in Table 1, we can specify several groups to which the given relationships fit. The first of them is defined by relationships 4 and 5. It is characterized by the lack of reciprocity, longevity and a very large difference in in-degree parameters describing two vertices in a pair. Moreover, it is worth noting that the vast majority of interactions take place in the evening and at night. So here we have an example of relationships, when one group of individuals who does not get enough in the social network by writing their own posts, communicates with very influential



individuals during the night hours. This is certainly a common behavior pattern that can be easily determined and used to predict the behavior of a given group of users in the future. Another group worth distinguishing are relationships 1, 6 and 7. They are characterized by a very comparable ratio of in-degree users involved in communication, there is a noticeable reciprocity of relationships and, what is important, a large single break in communication. The most likely interpretation of this phenomenon is the late establishment of relations between users, and then the relationship stabilization. This is also supported by high coefficients defining the longest sequences in which interactions occurred. Analyzing the relationship in terms of the time of establishing interaction, it is clearly visible that these relations, which are based on interactions at night, are most often one-sided relations. This thesis is supported by the relations 3.4 and 5. An interesting and very characteristic is the relations 12. This is a highlight of the results we have obtained, proving that the mutual relationship is favored by the situation when both sides share a similar social network. It is clearly visible that the bilateral relationship was established at the moment when the commenting person began to be more involved in the network and ceased to be anonymous.

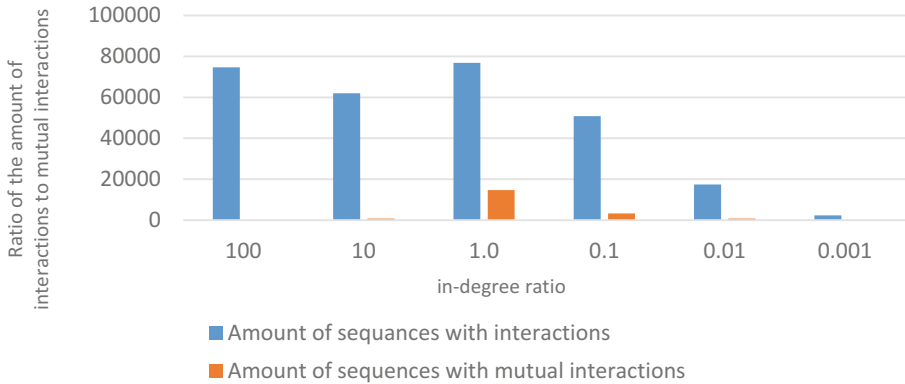
**Table 1.** Characteristics of the studied relationships, with the most frequent interactions.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	2	0.5	0.5–2	0.5–2	104	26.92	28	38	20	2	42	75	25	0	0
2	1.5	0.1	0.1–1.5	0.1–0.5	96	3.13	3	31	16	1	70	25	50	25	0
3	1.5	0.15	0.15–1.5	X	84	0	0	36	34	0	132	0	0	100	0
4	1500	500	500–1500	X	84	0	0	26	29	0	132	0	0	100	0
5	381	1500	100–1500	X	83	0	0	28	39	0	132	0	20	80	0
6	565	10	2–565	2–5	80	7.5	6	24	30	2	59	30	40	30	0
7	1	0.1	0.1–2	1	73	11.2	9	8	24	5	62	0	35	65	0
8	1.5	0.3	0.3–1.5	X	69	0.00	0	11	18	0	132	0	45	55	0
9	2	0.25	0.25–1	X	67	0.00	0	29	34	0	132	25	45	30	0
10	0.4	0.4	0.35–0.4	0.35–0.4	62	1.61	1	25	52	1	79	0	25	75	0
11	2	20	2–20	2–12	59	52.54	31	17	36	17	35	0	0	65	35
12	5000	10	10–5000	10–20	56	69.23	18	46	63	7	67	0	15	85	0
13	385	0.3	0.3–385	X	55	0	0	23	29	0	132	20	20	60	0
14	1	0.5	20	0.5–40	52	25	13	22	58	4	58	0	25	75	0

#### 4.6 The Influence of In-Degree Ration on Relationship

It should also be noted that the dominant factor that affects the development of the relationship between users is the ratio of their impact on the development of the network - the in-degree ratio. In Fig. 9 we show the ratio of the amount of interaction to mutual interactions depending on the in-degree ratio.

We note here that the highest chances of getting reciprocity in a relationship are those users whose in-degree parameter in the network is up to 10 times smaller than the commented person.



**Fig. 9.** Ratio of the amount of interaction to mutual interactions depending on the in-degree ratio (15-min slot).

#### 4.7 Experiments Conclusions

The goal set at the beginning of the work was to find relation patterns based on given attributes. At work we were able to analyze the relationships that occurred in a given social network. We presented the distribution of fast relations depending on the time window that the other side of the relationship had on reacting. As it turned out, increasing this window three times allowed to determine a 15-fold larger set of relationships. Similarly, we have done categorizing interactions depending on the time of the day, getting the answer to the information when there are the most, and therefore, when establishing communication gives the highest chance of response. An important piece of information for further analysis is that the vast majority of mutual relations occur where two parties with a similar social position in the network interact. We additionally presented specific relations, along with their characteristics, showing that with their help we are able to categorize a given relationship, and more importantly, and what is the subject of further work, predict how relations will behave in successive intervals. Our analyzes allowed to identify characteristics such as the frequency of interaction, the speed of interaction, their reciprocity and the time when they took place, which are key to determining patterns of relationships in the social network. Nowadays, with the increasing amount of data, it is increasingly difficult to build long-term relationships with many users and, as described in the article “The influence of their strength of importance in blogosphere”, these relationships are very often based on trust in the network. By using information about relationship patterns in a given social network, we are able to appropriately profile the activities of specific agents/users in order to achieve a specific goal.

## 5 Conclusion

In the paper we focus on the strong relationships between users in the blogosphere. The characteristics of such relations are analyzed considering on the social positions of the users, their duration and the periods of day with their highest intensity. The techniques

used to develop the model and the analysis of such modeled data, allow to effectively create social networks without the use of tools from external websites (such as API provided by social networking sites) while maintaining the privacy of users of these websites. Information on the time of activity and intensity of given user groups, while identifying these groups, allow for proper preparation of content provided by website owners. In addition, information about the probability of network development, maintaining mutual contacts between users, gives the opportunity to estimate traffic on the site. What's more, having information about the history of a particular user's relationship with other portal users, categorizing it, allows to present a specific offer for this particular user. Increasingly, social media offers different types of membership with more or less extensive functions delivered to users. It is worth noting that the benefits of using the mechanisms presented in this article may have the users of the portal themselves, by familiarizing themselves with the characteristics of specific groups, cliques, and profiling for a specific purpose of their own activity. In further work, we will be using data mining techniques to predict the duration of social relationships.

**Acknowledgments.** This work is partially funded by the Dean's Grant of the Faculty of Computer Science, Electronics and Telecommunications AGH UST.

## References

1. Bogdanov, P. Mongiovi, M., Singh, A.K.: Mining heavy subgraphs in time-evolving networks. In: 2011 IEEE 11th International Conference on Data Mining, pp. 81–90 (2011)
2. Borge-Holthoefer, J., Baños, R.A., González-Bailón, S., Moreno, Y.: Cascading behaviour in complex socio-technical networks. *J. Complex Netw.* **1**(1), 3–24 (2013)
3. Gliwa, B., Koźlak, J., Zygmunt, A., Cetnarowicz, K.: Models of social groups in blogosphere based on information about comment addressees and sentiments. In: Aberer, K., Flache, A., Jager, W., Liu, L., Tang, J., Guéret, C. (eds.) *SocInfo 2012*. LNCS, vol. 7710, pp. 475–488. Springer, Heidelberg (2012). [https://doi.org/10.1007/978-3-642-35386-4\\_35](https://doi.org/10.1007/978-3-642-35386-4_35)
4. Harush, U., Barzel, B.: Dynamic patterns of information flow in complex networks. *Nat. Commun.* **8**, 2181 (2017)
5. Hui, C., Tyshchuk, Y., Wallace, W.A., Magdon-Ismail, M., Goldberg, M.: Information cascades in social media in response to a crisis: a preliminary model and a case study. In: *Proceedings of the 21st International Conference on World Wide Web (WWW 2012 Companion)*, pp. 653–656. ACM, New York (2012)
6. Hulovatyy, H., Chen, T., Milenkovic, T.: Exploring the structure and function of temporal networks with dynamic graphlets. *Bioinformatics* **31** (2014). <https://doi.org/10.1093/bioinformatics/btv227>
7. Kabutoya, Y., Nishida, K., Fujimura, K.: Dynamic network motifs: evolutionary patterns of substructures in complex networks. In: Du, X., Fan, W., Wang, J., Peng, Z., Sharaf, Mohamed A. (eds.) *APWeb 2011*. LNCS, vol. 6612, pp. 321–326. Springer, Heidelberg (2011). [https://doi.org/10.1007/978-3-642-20291-9\\_33](https://doi.org/10.1007/978-3-642-20291-9_33)
8. Kovanen, L., Karsai, M., Kaski, K., Kertész, J., Saramäki, J.: IOP Publishing Lt, *J. Stat. Mech. Theory Exp.* **2011**, November 2011
9. Sekara, V., Stopczynski, A., Lehmann, S.: Fundamental structures of dynamic social networks. *PNAS* **113**, 9977–9982 (2016)

10. Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N., Hurst, M.: Patterns of cascading behavior in large blog graphs. In: Proceedings of the 2007 SIAM International Conference on Data Mining, pp. 551–556 (2007)
11. Milardo, R., Johnson, M., Huston, T.: Developing close relationships: changing patterns of interaction between pair members and social networks. *J. Pers. Soc. Psychol.* **44**(05), 964–976 (1983)
12. Morse, S., Gonzalez, M., Markuzon, N.: Persistent cascades: measuring fundamental communication structure in social networks. In: 2016 IEEE International Conference on Big Data, BigData 2016, Washington DC, USA, pp. 969–975, 5–8 December 2016
13. Oliwa, L., Kozlak, J.: Anomaly detection in dynamic social networks for identifying key events. In: 2017 International Conference on Behavioral, Economic, Socio-cultural Computing, BESC 2017, Krakow, Poland, 16–18 October 2017. IEEE (2017)
14. Pan, R.K., Saramäki, J.: Path lengths, correlations, and centrality in temporal networks. *Phys. Rev. E* **84**, 016105 (2011)
15. Rudek, K., Kozlak, J.: The influence of relationships strength on their duration in blogosphere. In: 2017 International Conference on Behavioral, Economic, Socio-cultural Computing, BESC 2017, Krakow, Poland, 16–18 October 2017. IEEE (2017)
16. Song, D., Wang, Y., Gao, X., Qu, S.X., Lai, Y.C., Wang, X.: Pattern formation and transition in complex networks, March 2017