CLUSTERING OF SOCIAL NETWORK USERS BASED ON TYPE OF THEIR ACTIVITY

V. Mishyn
National Technical University of Ukraine “Kyiv Polytechnic Institute”,
37, Peremohy Ave., Kyiv 03056 Ukraine
pti.mishin@gmail.com

Abstract. The purpose of this work is to offer clustering approach of social network users based on type of their activity, using automated monitoring information system in social networks, which is able to respond to an information changes and classify them. The proposed approach is specification that can be implemented in monitoring systems, which can be build for modern social networks. The approach does not claim to be accepted as canon, but rather reflects a basic analysis of public available information and shows how this information can be compared.

Keywords: social networks, classification, clustering, activity, account, posts, reposts, comments, likes.

Introduction. In the modern world social networking services become one of the key factors of the dynamics development of social relationships. There is no doubt that the impact of these networks on the flow of processes in the society is growing. For a long time receiving, distributing and processing information has become one of the leading industries. And with the development of social networks, each person has received more access and opportunity to become a source of dissemination of information, not only in the circle of his friends and in the vast expanses of the Internet. In the era of social networks, in which we now caught, is very important to have time to monitor the streams of information, because they describe events which had happened, or provoke events that will happen. And therefore it is necessary to pay attention to anomalies in the behavior of potential sources of harmful information.

Social Network Analysis provides an opportunity to observe the progress of any kind, especially social and civic events in the real environment. To conduct this analysis and observations is necessary qualified staff and software tools to automate the process. Ironically, in Ukraine there are not so many companies, which have started to make developments and researches in this area. And those who began to engage are choosing only commercial approach (mentioning brands in social networking groups, research of the brands audience, etc.).

In my opinion, this issue can’t be approached only in the viewpoint of getting profits. Field of work is so large and includes a large number of participants that in one world of social network there are people who exchange messages, video and informational content in general, and at the same time there is a communication between terrorist organizations which supervise their actions. Here question can be considered up to the level of national security.

Social networks now and public information
The social networking service is a Web site or other service in the Web that allows users to create public or half-public form, to make a list of users with whom they have a connection and browse a list of links and lists of other users.

The characteristic features of social networking are:
• Creating personal profiles, in which often need to specify the real and personal data, other information about yourself (study place and work place, hobbies, life principles, etc.);
• Providing of almost full range of opportunities for sharing information (location, photos, videos, text placement records (using blogs or micro-blogging), thematic groups, exchange personal messages, etc.);
• The opportunity to set and to support a list of other users with whom the user has some relationships (friendship, relationship, business and work relationships, etc.)

One of the most important functions of social networks is to provide actual information to its users. However this information inside the network is always influenced by many factors and can be significantly distorted or even be not correct from the start. Therefore the question of tracking way information streams in these social structures becomes extremely important in the viewpoint of providing users faithful and correct facts.
As it has become clear with such popularity of networks, and the growing number of users it is no longer possible and reasonable to spend human resource on looking through personal pages in network in order to find some information that may threat. So it is time to conduct the automation of this process. It is no wonder that now many different state structures, large corporations, small private firms and even some enthusiasts teams understand the importance of the role of social networks in shaping the information field, as a single man and society in whole, and develop a variety of tools for effective research and analysis of networks in the viewpoint of their impact on public opinion.

The experimental results indicate a significant influence of large social groups inside the network for current formation of public opinion. Just a minute from the time of the post in group of people who share it with their friends and acquaintances can grow to several hundreds or even thousands, and later to ten thousand of people. Of course these figures are much depending on the content of the post, its interest to the public, the onset, and maybe provocative.

Besides these per-second "news" can be forgotten as quickly as become popular, but still these scales and rates are impressive. In the case of sufficiently long, mass and concentrated of one subject with such groups means in networks may lead to quite substantial consequences. Each user of social networks that exist today has the ability to close information about him from extraneous users within the opportunities which provides to him service social network. But as strange as it is, many people do not pay the necessary attention for the issue of security of information about themselves.

Lots of people do not realize that the information which posted by them on social networks, can be found and used by anyone, not necessarily with good intentions. Information about members of social networks may find their employers, relatives, debt collectors, criminals and so on. Bailiffs sometimes use social networks to find defaulter or get information about their property.

For example, the social network "VK" if there is no restriction, in open access are available:

- General information
  - Name
  - Date of birth and age
  - Relationship Status
  - List of relatives
- User contacts
  - Country
  - Place of birth
  - Location
  - Mobile Phone
  - Home (optional) Phone
  - Skype
  - Link to personal website
  - Links to other social networks
- Users interests (from favorite books to activity)
  - User education
- Place and year of getting school education
- Career
  - Place and years of work
- Place and year of getting high education
- Position
- Military service
  - Country
  - Unit
- Life position
  - Political beliefs
  - Worldview
  - Personal priority
  - Important in others
- Views on smoking/alcohol
- Photos and videos albums
- Friends
- Groups and Pages from subscription

This is merely information that a user fills out for itself. More information can be found in groups, and in the pages of friends.

- Dialogues and debates with others
- Records on foreign pages
- "I liked" to posts

The most informative is the so-called "wall" user. It provides the following information:

- Notes of the user
- Repost posts of other users or groups
- Entries in the "wall" of other people
- Comments to posts
- "I liked" on the wall
And this possibility carries much more information about the person. According to these records, you can judge by the nature of man, in his interests and changing behavior.

Of course there are many users who understand that spreading so much information about them in open access is, at least, not wise. So they use the possibilities to sensitive information from extraneous users. Or do not fill the questionnaire. The most sensitive information - manual correspondence of course remains closed.

The number and variety of information that can be obtained in the open access is really impressive. Equally amazing is ignorance of people about the need to keep this information to themselves. Social networks themselves encourage to openness and informing people of unknown data itself. But this is also the benefit: a careful analysis can make a complete picture of who is a member of social network.

**Method of monitoring users to obtain and analyze information**

The automated monitoring system may be realized by using API, which provides a social network, or can be written, with the use of most modern programming languages. But it should be automated to receive all relevant information about the user at a time and keep track of changes in it. For this performed the monitoring process, which iteratively repeats the stages through the given intervals:

- Getting information from the Social Web
- Parsing data to store in the database
- Save a received data
- Comparison with previous results
- Signaling the appearance changes

Therefore, it is possible to receive notices about certain changes related with the user. For example, appearing new wall posts and comments, changes in questionnaire posts in certain groups or communities.

So, moving to classification: first you need to highlight the main ideas and understand what is needed to make the concept of activity. As mentioned earlier, the "wall" carries the most important information, and therefore in the context of the developed classification of users social networks we consider the activity of users - their behavior on their wall and the walls of others.

The main types of the activity on the “walls” in modern networks are listed in figure actions. Monitoring results for these actions can be classified according to its user's activity and compared with the activity of other users, or averaged notion of normal behavior. With a large deviation from the average standards of conduct user activity can be regarded as anomalous, for example:

![Fig. 1 Types of activity](image1.png)

![Fig. 2 Dynamics of posts depending on subscribers dynamics](image2.png)
This graph shows that in the sample, a certain group of people over which monitoring was performed behaved suspiciously. They have begun to increase their number of links without changing behavior. Or start trying to actively influence the circle of their ties, not expanding it.

![Graph showing the dependence of posts and time spent online](image)

**Fig. 3** Dependence of posts and time spent online

The behavior of people in selected with an ellipse area is anomalous. They spend so little time online in average but they make a lot of posts. Obviously, these users are spam bots.

You can also suggest dependence:
- Increase in post / repost / comments / likes from time spent online.
- Increase connections based on time spent online.
- The dependence of a selected type of the activity to another.

**Conclusion**

In the era of social networks it’s necessary to pay adequate attention to the research capabilities of automated information from them. For information spread through social networks and communities and pages in them do not concede, and are even faster than the spread of relevant content sites. Therefore it is important to notice the influence of agents as single users or entire groups. For the sake of convenience and in accordance with the activity they need to be classified. It helps find abnormalities of behavior, and people who are trying to influence the opinion of users in social networks on many topics and issues.

**References**
