Tie Activity Profiles in Social Networks

Vitaly Khudobakhshov
Odnoklassiniki LTD (ok.ru)
St.-Petersburg, Russia
vitaly.khudobakhshov@gmail.com

***Abstract*—This paper investigates properties of ties in large social networks in terms of user activity. In social networks people maintain communication with friends. It is supported by various types of interactions including sending private messages and likes, sharing various type of content etc. How many natural classes of social ties may exist in relation to user-to-user interaction? How these classes can be characterized? How many action types characterize one particular tie? This research reveals that social tie can be described by one dominant type of action. Some additional properties of social ties such as symmetry are also discussed.**

I. Introduction

Modern social networks form special area of research, but many aspects of social interactions are unavailable outside of the corporate research groups or/and restricted by API functionality.

Odnoklassniki ("classmates" a transliteration from Russian) is the second largest social network in Russia with more than 200 millions users and more than 12 billions edges in social graph. This is a huge data source which is generated by users. Many aspects of social communications can be measured in large scale in this rich data environment.

In the most cases an interest to social ties appears in the context of information flow and diffusion [1], [2], [3]. Analysis of social graph in terms of user interactions is provided in [6], [8], [9], [10], [11]. Relevant research in user behaviour can be found in [4], [5], [7] among many others.

In this paper we tried to find natural classification of social ties in terms of user-to-user interaction. This kind of research requires access to data unavailable via API e.g. message from one user to another. In many papers ties are considered in relation to their strength which is important for information diffusion, but here we would like to understand how tie can be characterized by corresponding actions (see section below for definitions). Common sense tells us that ties are characterized by many types of activities. However we would argue that this position is far from reality and that almost always ties may be defined by one specific type of action. In a sense our approach can be seen as a version of user behaviour analysis with certain restrictions.

In this respect few other questions naturaly arise. For our particular social network edge in social graph does not have a direction (friendship is a symmetric relation). But this symmetry is irrelevant if we are dealing with user-to-user interaction. What kind of ties still tend to be symmetric?

A. Definitions

Social ties can be viewed from various points of view. Generally tie is an edge (real or virtual) which can be considered as directed or undirected depending on situation. Some edges contain label which describes a type of relationship associated with this particular tie. In the most cases these labels are *love*, *spouse*, *parent* and other *relatives*, *classmate*, *colleague*, *coursemate* and *companion in arms*. On other hand a tie can be active or inactive in terms of number of actions associated with it. This actions form profile of the tie. Depending on functionality of social network the number of possible actions can considerably vary from tens to hundreds. But many of these actions are more frequent than others. For example *create message* and *like* are most frequent actions except some "silent" actions like viewing a photo.

Let us define main term of the paper carefully. *Social tie* here is an edge of undirected social graph with additional direction induced by type of interaction. For example one person can "like" content of another person and the response is a visit to the person's profile page. Futhermore social tie can exist without particular edge in the social graph (i.e. friendship relation).

We assume that labeled cluster defines a corresponing class with the same name, so we will use terms "cluster" and "class" interchangeably depending on the context.

B. Assumptions

Here we should mention some general assumptions about the social network and its users behaviour:

- Our social network is a general purpose one. It means that there is no natural bias to one particular kind of communication e.g. we are not just a messanger or video service. It also means that all kind of actions should be equally comfortable for users.
- From time to time a particular tie may change its class membership. Therefore we need to measure tie activity within not very long timeframe (for example, a month or two).

II. Main part

In this section experiment design and results are described.

A. Dataset properties

Our dataset includes one month activity of randomly choosen circa 400000 active ties. This is a result of aggregation

ISBN 978-952-68397-0-7 (paperback), ISBN 978-952-68397-1-4 (PDF)

of many billions rows of raw action data and subsequent filtering by the number of actions.

The first problem we have faced here was what kind of tie (in terms of the range of actions) should be considered as active. If we limit ourselves to the dataset with maximum 99 actions per month we will remove only 0.8 percent of the whole data. The main question is how we should define the minimum number of actions. In our case we decided to consider the ties with 6 and more actions.

Later we will discuss resulting cluster properties but now we just point out that this is a proper good choice we are to cover a reasonable set of ties.

The next step performed is a selection of action types which can potentially form a tie profile. For this purposes the most frequent actions will be choosen by selecting actions with the mean value > 0.05.

Here are the most important of the selected actions described:

- *create message* action is performed when a user has sent a *private* message to another one;
- *guest* action is performed when a user visits another user's profile; (who is not necessarily his friend). In this particular social network a user normally can see people who visit his profile page, except in case when an invisible mode was bought;
- *create photo comment* is performed when a user comments on another user's photo;
- *create status comment* is performed whan a user comments on another user's status;
- *like status*, *like photo comment* and *like photo* are "like" actions for the specified objects;
- *mark photo* is similar to *like photo* but have a more differentiated meaning;
- *send present* every user can send a *virtual* present to another one.

In order to profile a tie every action have been normalized by the total number of actions per tie. As a result we have had a dataset containing apart from tie identification (i.e. pair of user IDs) a column per action with the relative number of this particular action type.

B. Cluster analysis

According to the experiments with *k-means* method we found out that an optimal number of clusters is 6. Every cluster contains ties with certain characteristics. Let us describe our clusters: In the Table I actions were selected by the criteria where mean value is > 0.05 and the results are sorted by relative standard deviation:

$$Norm = SD/Mean$$

Despite the fact that every cluster has a second action the latter varies significiantly in terms of standard deviation. Therefore we suppose that every cluster can be described by its leading action component. The largest cluster is the cluster where primary action is sending message (#1). It covers about 30% of active ties. The second largest cluster is cluster *like photo* with 20%. The smallest one is the cluster where primary action is *like status* with 4%. Others are equally large clusters with about 15% each. An interesting phenomenon is the existence of separate cluster which is determined by *send present* action.

TABLE I. STRUCTURE OF CLUSTERS

Cluster	Action	Mean	SD	Norm
1	*create message*	0.80	0.17	0.21
	guest	0.08	0.10	1.27
2	*guest*	0.53	0.30	0.57
	create message	0.12	0.16	1.32
	like photo	0.07	0.12	1.71
3	*send present*	0.76	0.20	0.26
	create message	0.08	0.13	1.66
4	*like status*	0.78	0.22	0.28
	like photo	0.05	0.11	2.08
5	*mark photo*	0.73	0.22	0.31
	like photo	0.13	0.17	1.35
6	*like photo*	0.79	0.20	0.25
	guest	0.05	0.09	1.65

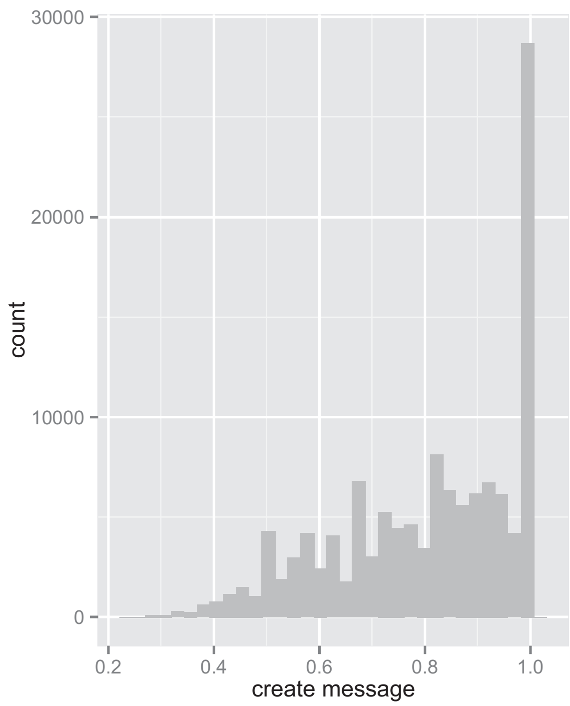

Fig. 1. Structure of *create message* cluster

In our initial experiments we tried to set various numbers of clusters. Let us say that k is a number of clusters. It was found that for $k < 6$ a structure of clusters was unstable. In some cases we ended up with clusters that had no particular structure: there was no action with $Norm < 1.0$, but in the most cases we had some of the clusters identical with the case when $k = 6$ plus one cluster without any structure. In this case the smallest cluster *like status* usually simply disappeared. For $k > 6$ we had basically the same situation. Clusters were stable only for $k = 6$.

Our cluster structure is almost identical as presented in Fig. 1. If we consider a case where fraction of *create message* actions in the whole set of actions equals 1.0 then the number of all actions distributed is shown on Fig. 1. *Guest* cluster is the only one with the large number of errors (about 7%). Despite its stability it contains a portion of records without the dominant action (number of action *guest* equals 0). Distribution of the dominant action is shown on Fig. 3. Nevertheless the peaks in 0.5 and 1.0 were distibuted in the same manner as on Fig. 2.

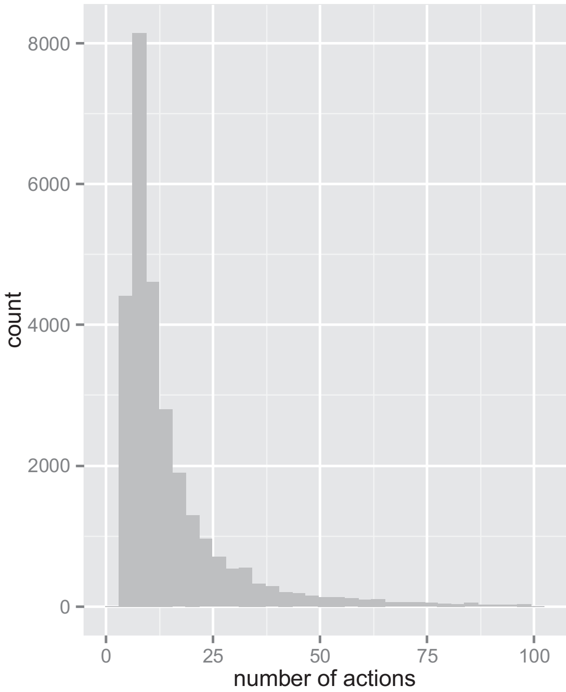

Fig. 2. Distibution for the rightmost bar on Fig. 1

Structure of clusters can be considered as trivial but it contradicted our initial intuition about the rich structure of social ties.

C. Classification

One can address this issue as a problem of classification. If we check our clusters with a classification algorithm it gives us a good evaluation method of the clusters. If we split our labeled dataset into train and test set as 2:1 and try to predict tie class (i.e. its membership to a particular cluster) we will have a confusion matrix with the extremely small amount of errors (presented in Table II). Overall accuracy of such classifier is

TABLE II. CONFUSION MATRIX FOR C4.5 CLASSIFIER

	1	2	3	4	5	6
1	41841	35	11	0	11	1
2	27	20016	10	8	35	17
3	10	20	20310	0	7	0
4	2	11	0	5730	1	2
5	4	14	2	1	15332	8
6	4	15	4	6	13	25936

0.997 (without removing errors in *guest* cluster).

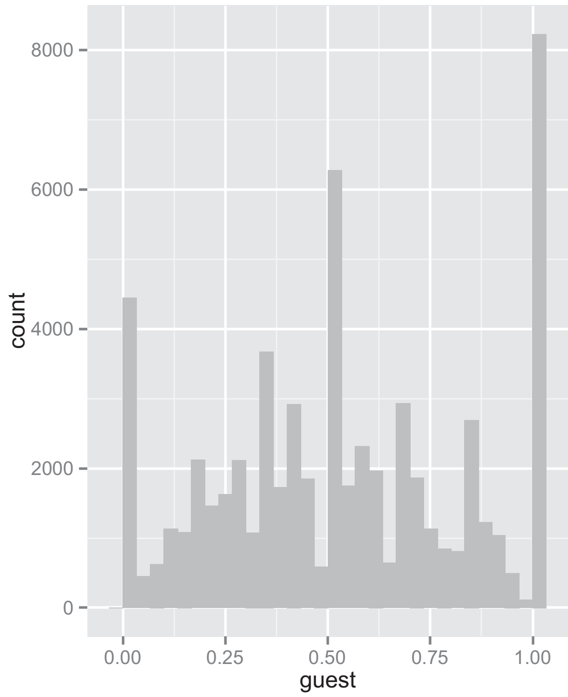

Fig. 3. Structure of *guest* cluster

D. Symmetry of ties

Friendship relation may be symmetric (like in Facebook) or asymmetric (like in Twitter) depending on intrinsic properties of the social network. Odnoklassiniki has symmetric friendship relation but in terms of activity associated with particular tie (in general friendship ralation is not needed for communication) we may have an asymmetric relation. Here we examine symmetry property of tie. In the Table III class-to-class symmetry matrix is presented. One can see that only *create message* and *send present* classes produce symmetric relations in the most cases.

TABLE III. SYMMETRY OF TIES

	1	2	3	4	5	6
1	93496	5517	3017	441	1686	2685
2	5517	7770	2095	551	1231	2536
3	3017	2095	24882	946	1381	2974
4	441	551	946	742	256	607
5	1686	1231	1381	256	2406	2815
6	2685	2536	2974	607	2815	6008

E. Discussion and future work

One of the most important issues we need to discuss is stability and polarity of clusters. Choosing minimum total number of actions per tie is a crucial thing. We restict a concept of active tie to 6 actions per month and more. If we increase this value to 20 actions we loose *like status* cluster (as it is then mixed with *guest* cluster) and *create message* cluster is splitted into 2 separate quasi-clusters (89% of the action and 51% respectively). This issues can be addressed by the assumption that users who spend more time in social network tend "to use" a particular tie in a more comprehensive way. All other classes were the same as in the first experiment. In all cases clusters tend to be "polar" (only one action characterizes a tie). Relationship between clusters' structure and user activity classes needs to be studied further. Here we only need to mention that the number of ties in the second experiment was much smaller compared with the first one.

Other questions to be addressed in the future is the existence of subclusters and tie dynamics.

III. Conclusion

The main result presented here can be described as follows: *a tie can be fully characterized by the most frequent action between corresponding users*. Moreover there is no class with two dominant actions. On other hand a friendship relation which is a symmetric one by definition and can be stratified by corresponing user behaviour.

This is the result which requires some psychological interpretation. Right now we do not know the real nature of this phenomena. It may be an intrinsic property of this particular social network or it can be true for the wide range of social media.

These results can be applied in many areas like information diffusion, customer behaviour analysis and link prediction.

Acknowledgment

Author would like to thank Dmitriy Bugaichenko, Vyacheslav Baranov and Viktor Ivanov for useful comments.

References

[1] Granovetter, Mark S. "The strength of weak ties." *American journal of sociology* (1973): 1360-1380.

[2] Krackhardt, David. "The strength of strong ties: The importance of philos in organizations." *Networks and organizations: Structure, form, and action* 216 (1992): 239.

[3] Bakshy, Eytan, et al. "The role of social networks in information diffusion." *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012.

[4] Benevenuto, Fabrcio, et al. "Characterizing user behavior in online social networks." *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*. ACM, 2009.

[5] Thelwall, Mike. "Social networks, gender, and friending: An analysis of MySpace member profiles." *Journal of the American Society for Information Science and Technology 59.8 (2008)*: 1321-1330.

[6] Wilson, Christo, et al. "User interactions in social networks and their implications." *Proceedings of the 4th ACM European conference on Computer systems*. ACM, 2009.

[7] Noulas, Anastasios, et al. "An Empirical Study of Geographic User Activity Patterns in Foursquare." *ICwSM* 11 (2011): 70-573.

[8] Viswanath, Bimal, et al. "On the evolution of user interaction in facebook." *Proceedings of the 2nd ACM workshop on Online social networks*. ACM, 2009.

[9] Huberman, Bernardo A., Daniel M. Romero, and Fang Wu. "Social networks that matter: Twitter under the microscope." *Available at SSRN 1313405* (2008).

[10] Golder, Scott A., Dennis M. Wilkinson, and Bernardo A. Huberman. "Rhythms of social interaction: Messaging within a massive online network." *Communities and technologies 2007*. Springer London, 2007. 41-66.

[11] Jiang, Jing, et al. "Understanding latent interactions in online social networks." *ACM Transactions on the Web (TWEB)* 7.4 (2013): 18.