Dynamics of Roles in the Context of Groups Evolution

Bogdan Gliwa, Anna Zygmunt, Jarosław Koźlak Department of Computer Science AGH University of Science and Technology, Poland {bgliwa,azygmunt,kozlak}@agh.edu.pl

Abstract—The paper addresses the analysis of a social system with identified groups and roles assigned to users. The important elements of this representation of a society and its dynamics are, on one hand, the identification of various important users in the whole network or in given groups, and, on the other hand, the events describing how the groups evolved. In this paper, we propose an approach integrating both these areas that would allow us to draw conclusions regarding the influence of important persons on the groups evolution and may make it possible to predict the future of the group on the basis of its current structure.

Keywords—roles dynamics; blogosphere; local roles.

I. INTRODUCTION

Taking into account a dynamic growth of different forms of social media, where users often express their opinions, it is important to understand the behaviour of users participating in them, identify their importance and predict directions of evolution. For such research, the application of methods of social network analysis became very popular. Considering the size and dynamics of changes, such a network may be perceived as a set of groups, within which users are more strongly connected than with others outside the group. Examples of such connections are discussions on forums or in blogosphere. The variety of subjects results in users belonging to many groups and participating in them with different levels of commitment, playing different roles in them.

Taking into consideration high dynamics of changes, an important question is, why some groups last for a long time and why others are more fugitive. It seems that the kinds of activities of the users within groups may have an influence on the duration of the groups. In this paper, we extend our algorithm for the analysis of the evolution of groups in blogosphere [1], [2] by including role identification in groups [3] and analysis of their evolution.

The organisation of the rest of the paper is as follows. In Section 2, the related works about group extraction and dynamics as well as identified roles of users are presented. Section 3 shows the model of the group dynamics, role identification and introduced R-SGCI method. In Section 4, the data set and performed experiments are described. Section 5 concludes the paper.

II. RELATED WORK

A. Groups extraction

One of the important problems in the analysis of social networks is the identification of the groups constituted by strongly connected nodes in the graph. Many algorithms were proposed so far and it is possible to organise them using different classification rules. One classification detects either non–overlapping or overlapping groups. The significant example of the algorithm which extracts ovelapping groups is the Clique Percolation Method (CPM), based on the identification of k-cliques [4]. The CPM algorithm is also used in the works presented in this paper. Other examples of overlapping approaches are solutions proposed by Shen et al. [5] and Gregory [6]. Examples of solutions working on non–ovelaping communities are algorithms proposed by Girvan and Newman [7] or Blondel et al. [8]. The wider overview of the group identification methods may be found in [9] [10].

B. Dynamics of groups

A typical way of analysing dynamic network is to divide the whole network into a series of static snapshots (called time slots or time steps). Greene et al. [11] presented a general strategy of analysis of dynamics of groups. Firstly, in each time slot the groups are extracted, and then groups from neighbouring slots are matched - it is performed by calculating Jaccard index between these groups and if the value of such a measure is above a predefined threshold, it means that the groups are matched. We [1] proposed Stable Group Changes Identification (SGCI) algorithm, which contains some improvements over earlier mentioned approaches. Instead of Jaccard index, we defined a new measure which has better properties in terms of matching groups with different sizees (Jaccard index is not well-suited for matching groups with significantly different sizes, because in such case the threshold should be very low and, at the same time, very low threshold applied to groups with very similar size could result in those groups being very big and the number of common members is very low, which is not the desired effect).

Another approach to analyse the dynamics of groups is presented in [12] and it is based on the use of CPM algorithm for group extraction. The first step is also finding communities in each time slot, but the process of matching groups from neighbouring time slots is quite different. For each consecutive time steps t and t + 1, the joint graph is built (the union of graphs from these 2 time steps). Next, in such joint graphs the groups are extracted and if a group from t time slot and a group from t + 1 time slot are contained inside the same group in the joint graph, it is assumed that these groups are matched.

Another important aspect is the identification of events that can occur in the group lifecycle. The set of events varies in different methods. Palla et al. [12] identified some basic events that can happen to a group: growth, contraction, merging, splitting, birth and death. Takaffoli et al. [13] used only 5 events: form, dissolve, survive, split and merge, but they additionally labelled transitions between groups as: size transition (group shrinks or group expands), compactness transition (group becomes more compact or more diffused), persistence transition (number of nodes and edges in group does not change), leader transition (when the node with the highest centrality in group does change).

C. Roles

One of the most popular definition of roles is that given by Wasserman in [14], where a role is identified as a position that has a distinct pattern of relations to other positions. Gleave in [15] distinguished two main methodological approaches to finding roles: interpretative and structural. Interpretative analyses employs methods such as etnography, content analysis, and surveys to capture behaviors and relations within groups. Structural analysis uses Social Network Analysis (SNA) and assumes that role entails a specific structural position. The most general approach to finding roles consists of two main stages [16]: understanding the community in order to identify potential roles, and then the creation of a role with observed characteristics and rules that will allow the classification of individuals into the pre-defined roles. One can distinguish several approaches for identifying social roles. The oldest one is based on equivalence classes [14], where the most appropriate is regular equivalence. Another approach is based on the identification of the core and periphery structure [17] where role is assigned based on membership of a particular area. In approach based on clustering feature vectors, each person is represented by a vector of some of the features that represents its behavior and relationships with the other members of the community and such vector can be clustered [16], so that people with similar characteristics are placed in one group.

III. MODEL

A. Dynamics of groups

A common approach to analyse dynamic networks is to divide the whole range of time into smaller periods (called *time slots* or *time steps*) and, then, in each time slot the static network is analysed.

For experiments we employed SGCI method [1] to analyse the dynamics of groups. The method is composed of four stages: identification of short-lived groups in each separated time interval, identification of group continuation, separation of the stable groups (lasting for a certain time interval) and the identification of types of group changes (transition between the states of the stable group).

Step 1. In each time slot the groups are extracted (such groups are called fugitive groups). Any method of group finding can be used for that purpose (in this paper we utilized the CPM method).

Step 2. In this step, the algorithm identifies transition between groups observed at time t and the groups observed at time t+1 (their successors). Identification is performed by calculating the Modified Jaccard Measure (A and B are examined groups from neighbouring time slots):

$$MJ(A,B) = \begin{cases} 0, & \text{if } A = \emptyset \lor B = \emptyset, \\ max(\frac{|A \cap B|}{|A|}, \frac{|A \cap B|}{|B|}), & \text{otherwise.} \end{cases}$$
(1)

and if the calculated value is above a defined threshold (in experiments we assumed the value equals 0.5) and the ratio of groups size

$$ds(A,B) = max(\frac{|A|}{|B|}, \frac{|B|}{|A|})$$
(2)

is no more than a specified value (in tests we used value equals 50), then group B is considered as continuation of the group A.

Step 3. In this step, the algorithm retrieves the stable groups. The stable groups are groups that exist in the required number of consecutive time slots – the groups which are not stable are rejected. In experiments such a number was equal to 3.

Step 4. Transitions between groups from neighbouring time steps are labelled by event names that describe a type of occuring change between groups. In the algorithm the following events are defined (A is the source group and B is the target group in analysed transition; sh and dh are defined thresholds which were set in experiments to values 10 and 0.05 respectively):

• **split** takes place when a group divides into some groups that do not differ considerably (in terms of group size) from the predecessor group:

$$\frac{|A|}{|B|} < sh,\tag{3}$$

• **deletion** occurs when a group disintegrates into some successor groups and in analysed transition successor group is much smaller than the predecessor group

$$\frac{A|}{B|} \ge sh,\tag{4}$$

• **merge** happens when many predecessor groups form a successor group in the next time slot and the former groups have size that do not differ significantly from the size of the successor group

$$\frac{|B|}{|A|} < sh,\tag{5}$$

• **addition** occurs when several groups from the previous time slot create a group in the next time slot and in analysed transition the origin group is significantly smaller than the successor group

$$\frac{|B|}{|A|} \ge sh,\tag{6}$$

- **decay** takes place when groups do not exist in the next time slot,
- **constancy** means simple transition with very small change of the group size

$$\frac{abs(|A| - |B|)}{|A|} \le dh,\tag{7}$$

 change_size – simple transition with significant change of group size.

B. Roles in groups

Users can play different roles on a global level and different ones in each of the group they belong to. In this paper we focus on roles defined on a local level – the level of group. The set of roles we use for analysis in this paper, was proposed by us in [3] and the roles were described there in detail.

We differentiated two kinds of influential people: (i) selfish ones who focus on building only their own position – they comment mostly in their own threads (ii) social ones who also take part in discussions started by other bloggers and comment on their posts.

The roles presented take into account responses from other users on the content the user writes (both in the form of posts and comments). To meet this assumption, we defined *Post* and *Comment Influence*.

Post Influence for author a has the following form (in this definition we use the notation c(X, cond) that means the number of elements in X that every element of X fulfills condition cond):

$$PostInf_a = 4 \cdot c(p_a, pr \ge A_p) + 2 \cdot c(p_a, pr \ge A_p/2) + c(p_a, pr \ge A_p/4) - 2 \cdot c(p_a, pr < 1)$$
(8)

where p_a – posts of author *a*; pr – number of comments for a given post excluding the author's comments in his own thread; $A_p = 10 \cdot groupDensity \cdot groupSize$

Comment Influence for author a is calculated in the following way (in this definition we use the notation w(cond) that returns 1 when the condition cond is satisfied, otherwise – 0):

$$ComInf_a = 4 \cdot w(r_a \ge 1.25) + 2 \cdot w(r_a \ge 1) +w(r_a \ge 0.75) - w(cr_a < D_c) - 2w(cr_a < D_c/2) -4 \cdot w(cr_a < D_c/4)$$
(9)

where r is the the number of received comments from other users divided by the number of written comments by given authors; cr is a number of reiceved comments from other users; $D_c = groupSize \cdot groupDensity$.

To define roles we need also another measure ComEgo which is a ratio between comments written in own threads and all comments written by a given user.

Using the above definitions we can describe the set of roles:

- 1) Influential User (infUser): PostInf > 2 and ComInf > 0
 - a) Selfish Influential User: ComEqo > 0.75
 - b) Social Influential User: ComEgo < 0.75
- 2) Influential Blogger (infBlog): PostInf > 2 and $ComInf \leq 0$
 - a) Selfish Influential Blogger: $ComEgo \ge 0.75$
 - b) Social Influential Blogger: ComEgo < 0.75
- 3) Influential Commentator (infComm): ComInf > 0and $PostInf \le 2$
- 4) Standard Commentator (comm): $c(comments) \ge 20$ and $c(posts) \le 2$
- 5) Not Active (notActive): c(posts) < 1 and c(comments) < 2

6) *Standard Blogger (stdBlog)*: User that does not match to any from above roles.

The above parameters were adjusted by us by testing with different values of them, comparing obtained results and finally verifying them using knowledge about bloggers.

C. Method of group dynamics analysis based on local roles

We introduced R-SGCI – modified version of SGCI algorithm that takes into account roles played by users in groups. It has additional condition enforcing passing influential roles between groups if they are present in the predecessor group:

$$RM(A,B) = \begin{cases} \frac{|R(A,Inf) \cap |A \cap B||}{R(A,Inf)}, & \text{if } R(A,Inf) > 0, \\ 1, & \text{otherwise.} \end{cases}$$
(10)

where R(A, Inf) is a number of users in group A with influential roles. In experiments, we rejected transitions between groups when the value of RM was equal to 0 - it means that if the predecessor group has any users with influential roles then at least one of them should be present in the successor group.

IV. DESCRIPTION OF EXPERIMENTS

A. Data set

The examined data set contains data collected from the portal salon24.pl, where discussions mostly concern political topics. The data set consists of 26 722 users (11 084 of them have their own blog), 285 532 posts and 4 173 457 comments within the period 1.01.2008 - 31.03.2012. The experiments described were carried out on half of this dataset - from 4.04.2010 to 31.03.2012. The whole period was divided into time slots, each lasting 7 days and neighboring slots overlap each other by 4 days. In the analysed period there are 182 time slots. In every slot the comments model was used, introduced by us in [18] - the users are nodes and relations between them are built in the following way: from user who wrote the comment to the user who was commented on or if the user whose comment was commented on is not explicitly referenced in the comment (by using @ and name of author of comment) the target of the relation is the author of post.

B. Groups and evolution events

In each time slot, the groups were extracted by CPM method – CPMd version from CFinder tool [19], for k equals 5. For group evolution we used the SGCI method (described in section III-A).

Fig. 1 shows the summary of groups with different sizes. As we can see, the most numerous part of all groups are the groups with the size equal to 5.

Fig. 2 presents the number of evolution events in the analysed data set. The most popular ones are additions and deletions, which occur when small groups attach to or detach from, respectively, a much larger group. As we could observe in Fig. 1, the smallest groups are the most numerous and due to their small size it is quite easy to find them matching with larger groups, so this could explain the huge number of these events.



Fig. 1. Summary of sizes of stable groups.



Fig. 2. Summary of events in analysed time slots.

C. Local roles

Table I shows how big a part of all users having a given role in group is present in a group in the consecutive time slot. The difference is most visible in the value of median for roles. During transitions in group evolution the people with the most important roles (Influential Bloggers and Influential Users) almost in every case are also active in the successor group.

 TABLE I.
 FRACTION OF USERS WITH GIVEN ROLE IN GROUP A THAT

 SUCH USERS PASS TO GROUP IN THE NEXT TIME SLOT.

role	mean	stdDev	median
infComm	0.648	0.395	0.778
comm	0.742	0.302	0.8
stdBlog	0.789	0.256	0.857
infBlogSel	0.85	0.336	1
infBlogSoc	0.885	0.305	1
infUserSel	0.832	0.34	1
infUserSoc	0.936	0.321	1

D. Local roles in transitions between groups

Fig. 3 shows the number and proportion of roles that users have (summed roles in all transitions) when they pass to any group in the next time slot. One can notice that Commentators and Standard Bloggers outnumber other roles. Influential roles constitute less than 10% of all roles.



Fig. 3. Local roles passing between groups.

Fig. 4 shows that different roles have different stability i.e. proportion of all cases that a user with a given role in a group will have the same role in a group in the consecutive time slot. The most stable roles are Commentator and Standard Blogger – they rarely become important users. Selfish roles have higher stability than social ones – selfish users only maintain their own threads. Moreover, Influential Blogger Selfish has higher stability than Influential User Selfish and Influential Blogger Social than Influential User Social. We can explain this situation that it takes more effort to play Influential User role than Influential Blogger (apart from writing influential posts, Influential Users have to also write influential comments), so it is harder to hold their roles.



Fig. 4. Stability of roles passing between groups.

Table II describes the transitions between roles i.e. number of occurences that a user with the first role (rows in the table) in a group was present in a group in the consecutive time slot and in that group the user has the second role (columns in the table). We can notice that Commentators and Standard bloggers mostly, after passing to another group, have the same role. Influential Commentators in the majority of cases also pass with the same role, but also quite large part has roles of Commentator or Standard Blogger (which have weaker conditions). Influential Blogger Selfish moves with the same role and also significant part of users with that role became Standard Bloggers in groups in the next time slot. Influential Blogger Social passes mostly to Standard Blogger role, but also a large part of them to Influential Blogger Social and Commentators. Influential User Selfish often becomes Standard Blogger, Influential Blogger Selfish or Influential User Selfish. Influential User Social proceeds to role Standard Blogger, and, less, to roles Influential Blogger Social and Influential User Social. In these transitions some more general observations can be formulated:

- most roles, except Commentators and Influential Commentators, pass in most cases to Standard Blogger
- social users, to a large degree, also transfer to social users and selfish users to selfish users
- Influential Users (Selfish and Social) proceed, in almost the same way, to Influential Bloggers and Influential Users, but Influential Bloggers only to Influential Bloggers.

E. Stability of local roles in evolution events

During experiments it seemed that the overall role stability for different events was very high and does not differ significantly. We looked into detail and it was caused by the fact that the most numerous roles are Commentator and Standard Blogger and they have very high stability for all events, as can be seen in Table III.

Table III presents how many times a user with a given role has the same role in any group in the next time slot for different evolution events and shows how often such a case took places in relation to all transitions of a given role in given event. We can observe that in each event the roles Commentator and Standard Blogger dominate over other ones (in terms of keeping their position). Influential Users frequently hold their role for simple transitions between groups – change size and constancy. Influential Bloggers often play the same role when the transition is one of following types: change size, constancy, merge and split.

F. Method of analysis of group dynamics with local roles

In this section we discuss results obtained with R-SGCI method. In Fig. 5, we can observe that the number of groups differs slightly between the original method (SGCI) and the modified one (R-SGCI). Differences are only in the number of small groups.

Fig. 6 presents the number of events that are obtained using both methods. The biggest difference is in the number of deletion events. It means that if a large group contains influential people and small group detaches from it (deletion event), then in most cases (around 2/3 of cases) the smaller group does not contain influential people.

In Table IV, there is a comparison of density and stability measure for groups acquired by both methods. Decreasing number of events (especially deletion event) and reducing number of small events explains increasing stability (weak events, such as deletion event, lower stability) and decreasing density (small groups are usually significantly more densier than larger ones).



Fig. 5. Comparison of quantity of groups between methods.



Fig. 6. Comparison of events between methods.

V. CONCLUSION

In the paper, the model of the social system with stable groups and roles is presented and a set of experiments was performed. The obtained results allow us to better understand behaviour of groups. The majority of the identified roles are less significant roles and we assume that only important roles (influential users, influential bloggers or influential commentators) have influence on group evolution. Generally, types of important roles (social, selfish) are preserved and passed to new groups in the next time period. One can also notice that a presence of influential roles significantly increases chances of groups lasting.

Future works may follow in several directions. The first is to analyse how the leaving of significant roles from groups influences the leaving of other group members. The second is an attempt to improve prediction methods taking into consideration important roles belonging to the group. We also plan to conduct experiments on other datasets.

Acknowledgments. The research leading to these results has received funding from the dean grant no. 15.11.230.083.

References

 B. Gliwa, S. Saganowski, A. Zygmunt, P. Bródka, P. Kazienko, and J. Kozlak, "Identification of group changes in blogosphere," in ASONAM 2012: IEEE/ACM International Conference on Advances in Social

TABLE II. NUMBER OF TRANSITION BETWEEN ROLES: SOURCE ROLES ARE IN ROWS, TARGET ROLES – IN CO	OLUMNS
--	--------

	infComm	comm	stdBlog	infBlogSel	infBlogSoc	infUserSel	infUserSoc
infComm	654	451	450	14	16	30	16
comm	407	22032	3100	106	157	9	15
stdBlog	483	3135	25772	751	559	104	80
infBlogSel	26	125	742	810	42	36	6
infBlogSoc	20	192	486	46	176	3	21
infUserSel	25	6	73	62	3	58	6
infUserSoc	15	18	75	13	21	11	20

TABLE III. FRACTION OF ALL CASES AND NUMBER OF OCCURENCES (AFTER SLASH) THAT USER WITH GIVEN ROLE HAS THE SAME ROLE IN THE CONSECUTIVE TIME SLOT FOR DIFFERENT EVOLUTION EVENTS.

	change size	constancy	merge	split	addition	deletion
infComm	0.491 / 340	0.554 / 104	0.505 / 74	0.339 / 89	0.514 / 21	0.061 / 26
comm	0.822 / 5862	0.852 / 2077	0.833 / 1806	0.859 / 1933	0.866 / 5087	0.859 / 5267
stdBlog	0.789 / 5559	0.813 / 2011	0.756 / 1489	0.83 / 1698	0.794 / 7392	0.893/7623
infBlogSel	0.607 / 201	0.612 / 67	0.722 / 179	0.613 / 161	0.368 / 101	0.211/101
infBlogSoc	0.3 / 56	0.36 / 18	0.375 / 35	0.245 / 34	0.103 / 14	0.056 / 19
infUserSel	0.392 / 30	0.517 / 15	0.222 / 2	0.152 / 4	0.375 / 3	0.046 / 4
infUserSoc	0.25 / 11	0.4 / 4	0 / 0	0.119 / 3	0.5 / 1	0.012 / 1

TABLE IV. COMPARISON OF MEASURES FOR GROUPS BETWEEN METHODS.

measure	SGCI [mean/stdDev]	R-SGCI [mean/stdDev]
stability	0.124/0.206	0.154/0.227
density	0.694/0.169	0.689/0.175

Networks Analysis and Mining: Istanbul, Turkey, ser. IEEE Computer Society, 2012, pp. 1233–1238.

- [2] B. Gliwa, P. Bródka, A. Zygmunt, S. Saganowski, P. Kazienko, and J. Kozlak, "Different approaches to community evolution prediction in blogosphere," in ASONAM 2013: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining: Niagara Falls, Turkey, 2013.
- [3] B. Gliwa, A. Zygmunt, and J. Kozlak, "Analysis of roles and groups in blogosphere," in *CORES - 8th International Conference on Computer Recognition Systems*, ser. Advances in Intelligent and Soft Computing, vol. 226. Springer, 2013, pp. 299–308.
- [4] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, pp. 814–818, 2005.
- [5] H. Shen, X. Cheng, K. Cai, and M.-B. Hu, "Detect overlapping and hierarchical community structure in networks," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 8, pp. 1706 – 1712, 2009.
- [6] S. Gregory, "A fast algorithm to find overlapping communities in networks," in *Machine Learning and Knowledge Discovery in Databases*, ser. Lecture Notes in Computer Science, W. Daelemans, B. Goethals, and K. Morik, Eds. Springer Berlin Heidelberg, 2008, vol. 5211, pp. 408–423.
- [7] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of The National Academy of Sciences*, vol. 99, pp. 7821–7826, 2002.
- [8] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, 2008.
- [9] A. Lancichinetti and S. Fortunato, "Community detection algorithms: A comparative analysis," *Phys. Rev. E*, vol. 80, p. 056117, Nov 2009. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.80.056117
- [10] L. Tang and H. Liu, Community Detection and Mining in Social Media. Morgan & Claypool Publishers, 2010.

- [11] D. Greene, D. Doyle, and P. Cunningham, "Tracking the evolution of communities in dynamic social networks," in *Proc. International Conference on Advances in Social Networks Analysis and Mining* (ASONAM'10), ser. IEEE Computer Society. IEEE, 2010, pp. 176–183.
- [12] G. Palla, A. Iszl Barabsi, T. Vicsek, and B. Hungary, "Quantifying social group evolution," *Nature*, vol. 446, pp. 664–667, 2007.
- [13] M. Takaffoli, J. Fagnan, F. Sangi, and O. Zaiane, "Tracking changes in dynamic information networks," in *Computational Aspects of Social Networks (CASoN), 2011 International Conference on*, 2011, pp. 94– 101.
- [14] S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge, 1994.
- [15] E. Gleave, H. Welser, T. Lento, and M. Smith, "A conceptual and operational definition of social role'in online community," in *System Sciences*, 2009. HICSS'09. 42nd Hawaii International Conference on. IEEE, 2009, pp. 1–11.
- [16] V. Junquero-Trabado and D. Dominguez-Sal, "Building a role search engine for social media," in *Proc. of the 21st Int. Conf. companion on World Wide Web*, ser. WWW '12 Companion. NY, USA: ACM, 2012, pp. 1051–1060.
- [17] S. P. Borgatti and M. G. Everett, "Models of core/periphery structures," *Social Networks*, vol. 21, no. 4, pp. 375 – 395, 2000.
- [18] B. Gliwa, J. Kozlak, A. Zygmunt, and K. Cetnarowicz, "Models of social groups in blogosphere based on information about comment addressees and sentiments," in *Social Informatics - 4th International Conference on Social Informatics, SocInfo, Lausanne, Switzerland*, ser. Lecture Notes in Computer Science, vol. 7710. Springer, 2012, pp. 475–488.
- [19] "Cfinder tool," www.cfinder.org, retrieved: 9,2013.