

GEVi: context-based graphical analysis of social group dynamics

Bogdan Gliwa · Anna Zygmunt

Received: 12 May 2013 / Revised: 13 November 2013 / Accepted: 30 December 2013 / Published online: 18 February 2014
© Springer-Verlag Wien 2014

Abstract Identifying communities and analysing their dynamics in social networks is very important research problem. However, qualitative analysis (taking into account the scale of the problem) still poses serious problems. Several methods for analysis are proposed, but there is missing tool allowing visualisation of dynamics of communities and enabling performing analysis on different levels of details. This paper describes a tool enabling analysis of social group dynamics with taking into account many aspects of groups (contexts). In paper the analysis of group density, sentiment and topic modelling for groups is presented. Presented results are based on real-world data from blogosphere.

Keywords Social networks analysis · Complex networks · Group identification · Group evolution · Dynamics analysis · Graphical analysis

1 Introduction

Current trends in identification of groups in complex network analysis tend to go beyond static analysis (see, e.g., Palla et al. 2007; Spiliopoulou 2011) and take into account the dynamic character of the environment, mostly concerning the quantitative analysis of such dynamic groups. Qualitative analysis becomes a very difficult task, due to huge network sizes, possible number of groups and time-

dependence. In this paper, GEVi (Group Evolution Visualisation)—a tool for the graphical analysis of the evolution of groups will be presented.

Real-life networks are characterized by rapid changes and the groups that may be located are mostly short-lived and elusive. In order to analyse certain processes or trends occurring in groups, different time periods should be taken into account. Observation of changes should lead into stating the reasons for creation, extension or disappearance of certain groups. It is to note, that an additional challenge is the fact, that one user may be a member of many groups. Correlating of the observation of the network dynamics with external events may lead to explaining of certain processes occurring in the structure of groups and to allow prediction of future events.

In the paper, after presenting the state of the art and describing the utilized method of groups extraction, the features of the presented tool are shown. An earlier version of the tool was described in (Gliwa et al. 2012). Its capabilities were there illustrated on the basis of one of the most popular data set: Enron emails.¹

2 Overview of research

2.1 Groups extraction

The existence of groups (often called communities) in social network is intuitively obvious (Porter et al. 2009) and has been studied for a long time, especially in sociology and anthropology. Initially finding groups in large social networks was made possible by extracting certain features from the network and analyze them on higher level

B. Gliwa (✉) · A. Zygmunt
Department of Computer Science, AGH University of Science and Technology, 30-059 Kraków, Poland
e-mail: bgliwa@agh.edu.pl

A. Zygmunt
e-mail: azygmunt@agh.edu.pl

¹ <http://www.isi.edu/~adibi/Enron/Enron.htm>

of abstraction: the network could be represented in an equivalent, but much less complex form as groups and the relationships between them (Wasserman and Faust 1994). Nowadays, group finding techniques allow not only to simplify the network, but moreover, to analyze certain processes in micro and macro scale. There are many definitions of a group (Wasserman and Faust 1994; Agarwal and Liu 2009; Evans and Lambiotte 2009; Fortunato 2010), but usually it is assumed that the group is a set of vertices which communicate to each other more frequently than with vertices outside the group. Many methods of finding groups—overlapping or not—(mainly in static graphs) have been proposed (Fortunato 2010).

Every group can be described by several parameters, e.g. density (ratio of the number of links within the group to the maximum possible number of links), cohesion (ratio of the average strength of links between the members to the average strength of their links with people outside the group) or stability between groups (the ratio of the number of people, present in both group to the number of all group members).

2.2 Groups life cycle

Nowadays, many results regarding the dynamics of the network, taking into account the time and its impact on the life cycle of the groups are published (Asur et al. 2009; Spiliopoulou 2011).

For dynamic network analysis the common way is to divide given period of time into smaller units called time slots. Then, in each time slot the static network is analyzed and the groups are extracted. Next step is to determine the transitions between groups from neighboring time slots. For this purpose, Greene et al. (2010) used the Jaccard index as a measure describing the similarity of groups (the measure is calculated for each pair of group from neighboring time slots). The value of this measure above arbitrarily defined threshold level means that one group is continuation of another. Some other measures for obtaining transitions between groups have been proposed in literature (Gliwa et al. 2012; Bródka et al. 2013).

Palla et al. in (2007) identified basic events (transitions) that may occur in the life cycle of the group: growth, merging, birth, construction, splitting and death. They did not give any additional conditions. Asur in (2009) introduced formal definitions of five critical events. Greene in (2010) presented a review of the fundamental events describing group evolution and formulated these key events in terms of rules.

2.3 Graphical presentation of groups evolution

Despite the importance of issue of group evolution, currently there are very few methods to visualize group dynamics and they neglect events from groups lifecycle.

In Reda et al. (2011), a tool for visualization of the evolution for non-overlapping groups was proposed. With this tool one can analyse the membership of certain individuals in the group, rather than the evolution of the group itself. The tool is focused on visualisation of migration between users in disjoint groups.

Federico et al. (2012) have introduced ViENA (Visual Enterprise Network Analytics), tool for observing changes in centrality at different levels: juxtaposition view, superimposition and a two-and-a-half-dimensional view. Network at each time interval is shown in a separate window in a layout, with the possibility of setting up a suitable coloring of observed nodes (in their paper the usecase with coloring groups was not presented, but the tool enables colouring nodes in some ways, so it may be used to visualise communities by different colours).

Beiro et al. (2010) developed SnailVis, tool for visualising disjoint groups in different time steps. In each time step groups are visualised as circles with their radius proportional to the number of internal connections and thickness of edges between groups shows number of links between nodes from these groups. The tool doesn't enable analysis of events of groups lifecycle.

In the article we are introducing a new visualisation method of groups evolution (in our tool GEVi). This method presents the dynamics of groups in the form of graph. Vertices of such graph represent groups from various time slots and edges indicate which group is continuation of another. The visualisation shows earlier mentioned events from group evolution including events described by us.

Table 1 presents comparison of features of available tools for the visualisation of groups dynamics. As we can notice, most of them does not support overlapping groups which often better describe relationships of users in networks, especially in social media networks.

3 The method of groups extraction in dynamic environment

We have used SGCI (Stable Group Changes Identification) algorithm (Gliwa et al. 2012) and CPM (Cliques Percolation Method) (Palla et al. 2005) as a group extraction method. The algorithm consists of four main steps: identification of short-lived groups in each separated time interval; identification of group continuation (using modified Jaccard measure, see formula 6), 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).

Modification of the Jaccard measure, corrects some of its drawbacks (e.g. high threshold for this measure causes

that the groups have to be of similar size to be treated as one of them is the continuation of the second one). Our modification removes this limitation by considering the common elements from both groups in each group (and at least in one of the groups the common elements should be above defined threshold) instead of taking into account the common elements in sum of both groups.

A detailed description of the algorithm is in (Gliwa et al. 2012) [and its previous version in (Zygmunt et al. 2012)].

We used the set of events identified in (Gliwa et al. 2012), applying more general methods for their identification. We expanded list of possible events (described in Sect. 2.1) with few complex cases which occur frequently in the data. The algorithm identifies transitions between groups observed at time t and the groups observed at the time $t + 1$ (their successors). This is achieved by comparing the size of the source groups, with each of their successors, rather than the difference in size between all successors.

Eight types of changes (transitions) were identified (Fig. 1):

- **split** occurs when group falls into several successor groups; the group, that the transition comes to, cannot differ significantly from the largest of successor groups (if it is the largest group, the transition is treated as simple transition—constancy or change size respectively),
- **deletion** means that group disintegrates into many successor groups and the successor group of this transition is significantly smaller than the largest group from the successor groups,
- **merge** when transition is one among few, which create a group in the next time slot, the size of the former group cannot differ significantly from the largest of predecessor group for the group that is created in the next time slot (if it is the largest group, the transition is treated as simple transition—constancy or change size respectively),
- **addition** when the given transition is one among several which create a group in the next time slot, the origin group for this transition is significantly smaller from the largest of origin groups,
- **split_merge** in the same time, a split of the original group and the joining of many groups into successor

groups took place, this transition is labeled as `split_merge` if the addition is not assigned earlier (we consider that the addition has higher priority),

- **decay** the total disintegration of the group, which does not exist in the next time slot,
- **constancy** means simple transition without significant change of the group size,
- **change_size** simple transition with the change of the group size.

For various reasons, it is interesting to observe lifespan of communities. How social network is evolving? Is it possible to find some rules, principles, and develop models that explain its evolution? What are the reasons for appearance of communities in social network, how they grow or shrink, what are the causes of new members joining and abandoning the old? Why sometimes the changes are smooth, and other times very rapid? Whether the community observed in two time periods is the same community, even though, for example, there is no common members? How change the character of communities, when new members come or old become inactive?

There are many interesting questions, but the available tools lack possibilities of simple, preferably graphical, analysis of groups life-cycle. A tool that may be used both for quantitative and qualitative analysis presenting graphical visualization of events and changes in the network would be much desired.

It would be simpler to visualize how the groups changed in response to some external events.

4 Model of social network dynamics

In this section a simple model for describing the analysis of the network dynamics is proposed.

4.1 General model

A complex network or social network may be of course described using standard definition of a graph:

$$N = \langle V, E \rangle \tag{1}$$

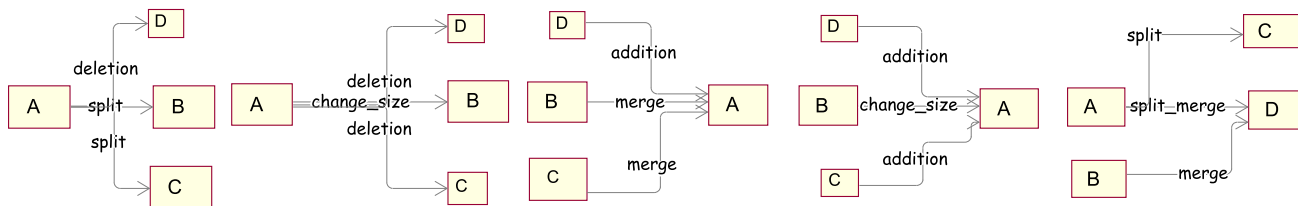


Fig. 1 Illustration of events

where: $V \subset \mathbb{N}$, stands for a finite set of vertices, that is:

$$V = \{i : i \in \mathbb{N} \wedge i < i_{\max}\} \tag{2}$$

and $E \subset V \times V$ is a finite set of edges.

Striving to provide means for observation of groups that are formed in a certain time moment, let us consider the following space of system states: $G = 2^V$. The elements of G are any possible subsets of V . Now, observing the system in a certain time moment, it may be seen that the set of vertices is decomposed into following subsets:

$$G \ni g_t = \{g_{t,k}\}, t, k \in \mathbb{N}. \tag{3}$$

each subset may be described as:

$$g_{t,k} = \{v_1, \dots, v_{\max_{t,k}}\}. \tag{4}$$

where $\max_{t,k}$ stands for maximum number of the individuals in the group. Note, that the subsets (later called as groups) observed at certain time t may contain the same elements (they may overlap).

4.2 Dynamics of social network

Now, let us define the graph depicting the dynamics of the complex network. Again, as it is a graph, the definition is similar to the classical one:

$$D = \langle V_D, E_D \rangle \tag{5}$$

where: $V_D = (t, k) \in \mathbb{N} \times \mathbb{N}$, and $E_D = V_D \times V_D$ so this graph is composed of labels utilized before, in the definition of the complex network and the groups. Note, that this definition spans to the whole observation time of the network.

The above-presented simple formalism is aimed to ease the definition of observed events and other primitives.

For example, let us define Modified Jaccard measure

$$MJ(A, B) = \begin{cases} 0, & \text{if } A = \emptyset \vee B = \emptyset, \\ \max\left(\frac{|A \cap B|}{|A|}, \frac{|A \cap B|}{|B|}\right), & \text{otherwise.} \end{cases} \tag{6}$$

and ratio of groups size

$$ds(A, B) = \max\left(\frac{|A|}{|B|}, \frac{|B|}{|A|}\right) \tag{7}$$

where $A \neq \emptyset \wedge B \neq \emptyset$.

Transition $t_{g_{i,k}, g_{i+1,l}}$ can be defined as:

$$t_{g_{i,k}, g_{i+1,l}} : \exists g_{i,k} \wedge \exists g_{i+1,l} \wedge MJ(g_{i,k}, g_{i+1,l}) \geq th \wedge ds(g_{i,k}, g_{i+1,l}) < mh \tag{8}$$

where th means minimum threshold for creation of transition (in experiments we set value of th to 0.5) and mh

means maximum allowed difference of group sizes (in experiment we set value of mh to 50).

Now we can label transitions (Fig. 1 shows illustration for most events):

- addition:
$$t_{g_{i,k}, g_{i+1,l}} : |g_{i+1,l}| / |g_{i,k}| \geq sh \tag{9}$$

- deletion:
$$t_{g_{i,k}, g_{i+1,l}} : |g_{i,k}| / |g_{i+1,l}| \geq sh \tag{10}$$

- merge:
$$\begin{aligned} t_{g_{i,k}, g_{i+1,l}} : & ds(g_{i,k}, g_{i+1,l}) < sh \wedge \\ [\exists t_{g_{i,m}, g_{i+1,l}} : & m \neq k \wedge ds(g_{i,m}, g_{i+1,l}) < sh] \wedge \\ [\exists t_{g_{i,k}, g_{i+1,n}} : & n \neq l \wedge ds(g_{i,k}, g_{i+1,n}) < sh] \end{aligned} \tag{11}$$

- split: occurs when group divides into 2 or more groups in next time slot and these groups from next time slot have similar size to the group that divides

$$\begin{aligned} t_{g_{i,k}, g_{i+1,l}} : & ds(g_{i,k}, g_{i+1,l}) < sh \wedge \\ [\exists t_{g_{i,k}, g_{i+1,n}} : & n \neq l \wedge ds(g_{i,k}, g_{i+1,n}) < sh] \wedge \\ [\exists t_{g_{i,m}, g_{i+1,l}} : & m \neq k \wedge ds(g_{i,m}, g_{i+1,l}) < sh] \end{aligned} \tag{12}$$

- split_merge: occurs when group $g_{i,k}$ divides into 2 or more groups in next time slot, these groups from next time slot have similar size to $g_{i,k}$, the group $g_{i+1,l}$ is created from 2 or more groups from previous time slot and these groups from previous time slot have similar size to $g_{i+1,l}$

$$\begin{aligned} t_{g_{i,k}, g_{i+1,l}} : & ds(g_{i,k}, g_{i+1,l}) < sh \wedge \\ [\exists t_{g_{i,m}, g_{i+1,l}} : & m \neq k \wedge ds(g_{i,m}, g_{i+1,l}) < sh] \wedge \\ [\exists t_{g_{i,k}, g_{i+1,n}} : & n \neq l \wedge ds(g_{i,k}, g_{i+1,n}) < sh] \end{aligned} \tag{13}$$

- constancy:
$$\begin{aligned} t_{g_{i,k}, g_{i+1,l}} : & \text{abs}(|g_{i,k}| - |g_{i+1,l}|) / |g_{i,k}| \leq dh \wedge \\ [\exists t_{g_{i,m}, g_{i+1,l}} : & m \neq k \wedge ds(g_{i,m}, g_{i+1,l}) < sh] \wedge \\ [\exists t_{g_{i,k}, g_{i+1,n}} : & n \neq l \wedge ds(g_{i,k}, g_{i+1,n}) < sh] \end{aligned} \tag{14}$$

- change_size:
$$\begin{aligned} t_{g_{i,k}, g_{i+1,l}} : & \text{abs}(|g_{i,k}| - |g_{i+1,l}|) / |g_{i,k}| > dh \wedge \\ [\exists t_{g_{i,m}, g_{i+1,l}} : & m \neq k \wedge ds(g_{i,m}, g_{i+1,l}) < sh] \wedge \\ [\exists t_{g_{i,k}, g_{i+1,n}} : & n \neq l \wedge ds(g_{i,k}, g_{i+1,n}) < sh] \end{aligned} \tag{15}$$

- decay:
$$\exists t_{g_{i,k}, g_{i+1,l}} \tag{16}$$

In above definitions we used function abs which means absolute value function and some parameters: sh threshold for ratio of groups size and dh threshold for groups size differences. In experiments we set value of sh to 10 and value of dh to 0.05.

4.3 Contexts

Context C_A represents one aspect of system e.g. one context is theme of discussion, another can be whether people talk in a positive, neutral or negative way (sentiment).

Each context C_A has some categories (subcontexts):

$$C_A = \{C_A^{ct_1}, \dots, C_A^{ct_N}\}. \quad (17)$$

Referring to example with *sentiment* as a context—there are 3 categories: *negative*, *neutral* and *positive*. For *theme* in discussion as a context the categories are sets of similar subjects of discussion (called later *topics*) e.g. *Politics*, *Sport*, *Education* etc. Both mentioned contexts—topics and sentiment are in the limelight of research (Macskassy 2011; Mostafa 2013).

We can define function cf that for each group $g_{t,k}$ and category $C_A^{ct_x}$ in context C_A assigns value:

$$cf(g_{t,k}, C_A^{ct_x}) = d \quad (18)$$

where $d \in [0, 1]$.

It's worth mentioning that concept of context is present in literature. Jung in (2011) described a concept of context for users and groups in a collaborative network (where users cooperate on their tasks) and he defined context as a set of concepts that match personal ontology of user (his knowledge about world) and resource which the user is working on. But we want to emphasize the difference between his and our definition—we treat context as an aspect (projection from multiple points of view) to look on groups or individuals. However, context in both approaches for groups can be calculated as a sum of contexts for individuals (in this paper we consider it only on group level and our visualisation regards this level).

5 Component tool for analysis of complex networks (COMET)

The COMET (COMplex network Exploration Toolkit) (Fig. 2) is a tool for analysing complex networks, especially social networks. The tool is built based on Eclipse 4 RCP platform² and contains many plugins related with analysis and visualisation of different aspects of networks. It uses graph database Neo4j³ as datastore. One of main advantages this tool is support for analysis of dynamics of networks. The analysed network can be divided into time slots (overlapping or disjunctive) and each time slot can be visualised as network. The tool can calculate many well-known SNA measures such as betweenness, closeness, PageRank or density (Wasserman and Faust 1994).

² <http://www.eclipse.org/home/categories/rcp.php>

³ <http://www.neo4j.org/>

Furthermore, COMET contains some algorithms of group extraction [Blondel et al. (2008), Edge Betweenness (Girvan and Newman 2002) and CPM (Palla et al. 2005] with using Cfinder⁴) and role calculation [Structural equivalence (Hanneman and Riddle 2005) and CATREGGE (Borgatti and Everett 1993)]. The COMET has plugin architecture and can be easily extended.

6 Tool for graphical analysis of network evolution (GEVi)

The GEVi visualizes groups in time slots and displays transitions between them in a form of graph. Each distinct hierarchy of group evolution is displayed as a separate graph. To implement visualisation we used JGraph⁵ (Java-based library). GEVi is a plugin in COMET tool (Fig. 3), but GEVi can be also used separately (as a library). Furthermore, the tool can also display evolution events between groups in the form of table (Fig. 4)—GEVi integrates SGCI (Gliwa et al. 2012) method for purpose of identification group evolution events (but can be extended for other methods of group event identification).

6.1 Visualisation technique

The groups and transitions between them are represented using hierarchical (Sugiyama type) layout. It (Bastert and Matuszewski 2001) has several interesting features: there are few edge crossings, the nodes are evenly distributed and the edges are as straight as possible. The Sugiyama layout is a method for visualizing directed graphs and consists of the following stages:

- **cycle removal** some edges are reversed in order to make the graph acyclic (at the end of algorithm they are reversed again to initial state),
- **layer assignment** assignment of the vertices to layers (if there are edges that pass not only through adjacent layers, the dummy vertices are introduced),
- **crossing reduction** in each layer the ordering of vertices is calculated in order to minimize the number of edge crossing,
- **coordinate assignment** positioning of vertices so they do not overlap each other and that vertices not lie on the straight lines between two adjacent vertices from different layers, placing edges.

In our case, the transitions between groups cannot form cycles in graph so we omitted first stage. The second stage was simple in our situation because the groups are assigned

⁴ <http://www.cfinder.org/>

⁵ <http://www.jgraph.com/jgraph.html>

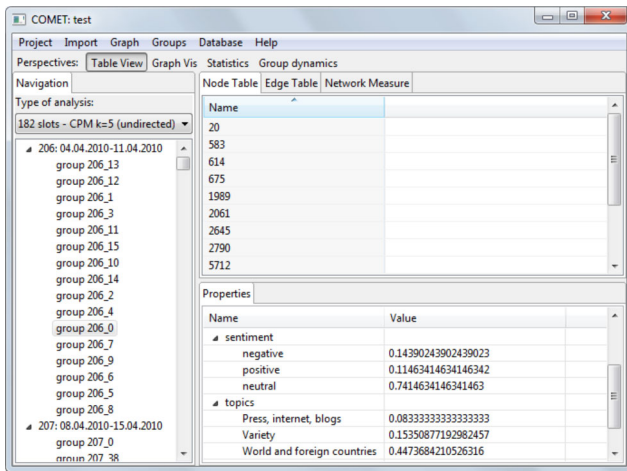


Fig. 2 COMET tool

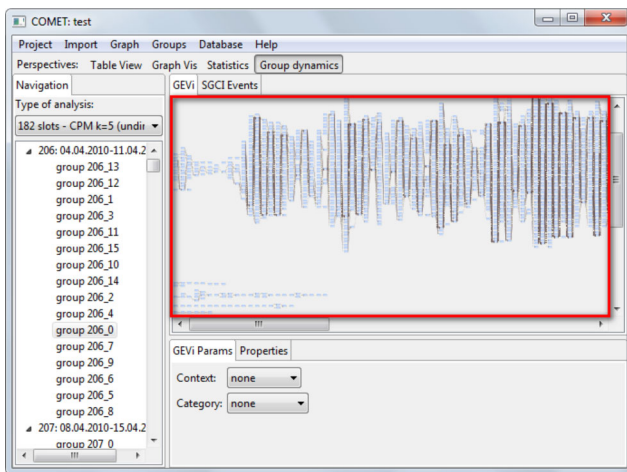


Fig. 3 GEVi plugin in COMET

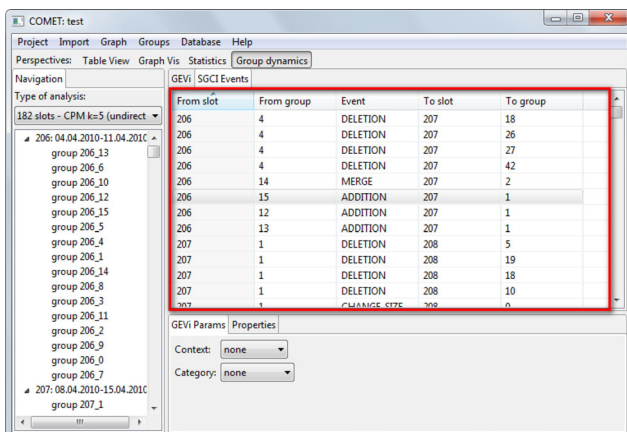


Fig. 4 GEVi table with group evolution events

to time slots where they were extracted. As the layers in the graph represent the time slots, so we preassigned nodes in the graph to their layers. For reduction of crossings and coordinate assignment, some variants of median method described by Gansner et al. (1993) were used.

6.2 Basic features

In GEVi, each group is labeled in a form *timeslotNumber_groupNumber* which eases the identification of the groups during their evolution. GEVi enables not only analysis of transitions between groups in different time slots (Fig. 5) but also shows the size of groups (in square brackets inside vertices), denoting how many members get inside the group during each group transition (label on transition) and how many of them get outside during each group transitions (in a form of number close to the green arrow—the green arrow pointing in the direction of the top-right corner stands for the number of members that go outside groups connected by outgoing transitions and the green arrow pointing in the direction of the bottom-right corner stands for the number of members that go into given group). For instance, the group 311_7 from Fig. 5 has 1 input edge (3 members flow from predecessor of that group to the given one) and additionally 2 members (not belonging to predecessor of that group) come to this group. The group has 1 outgoing edge (3 members flow to its successor) and additionally 2 members leave that group.

Some transitions are displayed as dashed arrows (Fig. 6)—this indicates that groups between given transition differ significantly in size (one of them is at least 10 times bigger than the second one). Such transitions represent events described as addition or deletion (depending whether small group attaches to the larger or small group detaches from the larger one).

In the transition pop-up menu, there is an additional information about stability (and event name) during group transition (Fig. 6) and in the group pop-up menu (Fig. 5)—the members of the group and intensities in contexts are listed.

GEVi also gives us information about overlapping of the members between the groups. After selecting of the group, all other groups that have in common at least one member with the selected one are highlighted (Fig. 7) and the information is displayed, regarding the number of common members (number between characters < and > inside vertex) and in the pop-up menu the members of all highlighted groups common with the selected one are shown.

To be more useful, GEVi supports also zooming graphs and searching for groups by its name in a form of *timeslotNumber_groupNumber* (after finding the group, the focus is set and the view is centered—Fig. 8).

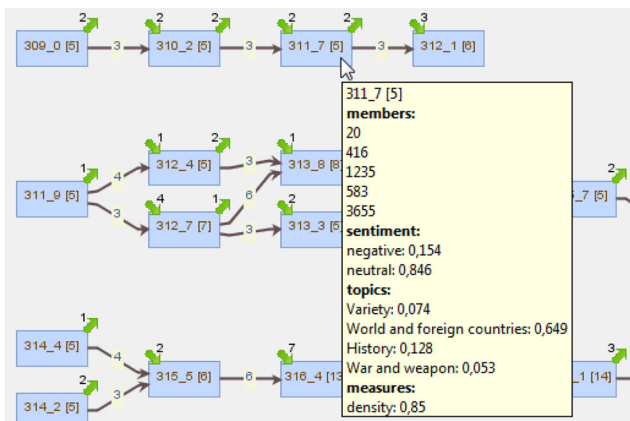


Fig. 5 Visualisation with showing context menu for group

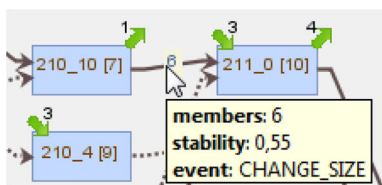


Fig. 6 Stability for chosen transition on visualisation

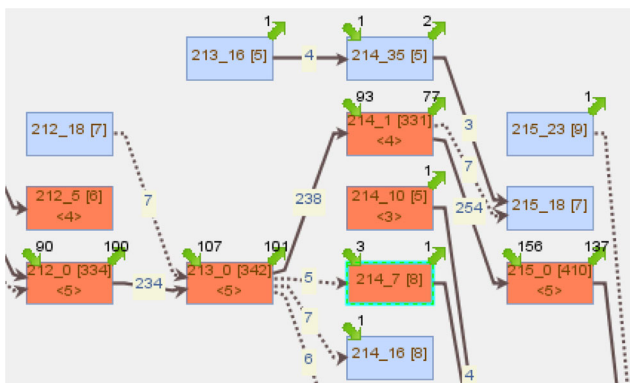


Fig. 7 Visualisation showing common members for group 214_7

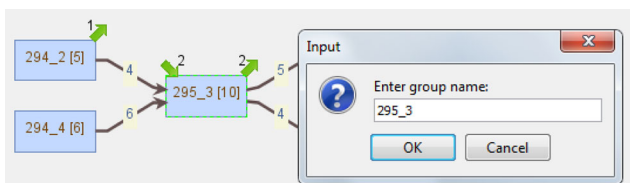


Fig. 8 Search for specific group on visualisation

6.3 Context related features

With each group we can associate context, category (within context) and value. Context means different analysed aspect such as topics discussed in groups, sentiment (what emotions are caused in people engaged in conversation within groups), even measures for groups can be perceived as context. Each context can have possibly numerous categories e.g. in sentiment context the categories are: positive, negative and neutral. For given context and category the node in GEVi (representing group) is coloured according to its value (called *lateness intensity*) in this context and category. We used color palette changing from blue (when the value is very low) to red (for high value). Threshold for value when red color should be applied is adjustable and defined by user (because context can have one or more categories and in some applications value 0.4 means high value and in other—low one). Therefore, in experiment for topics we set threshold equals 0.3 (values greater or equal 0.3 are coloured as red), for sentiment we set value 0.5 and for density—value 1.0. Figure 9 presents intensity of chosen category in topics context for groups.

7 Overview of graphical analysis on the basis of data from the blogosphere

In the article (Gliwa et al. 2012) the capabilities of a preliminary version of the GEVi were presented based on the Enron data set. It is relatively simple and not very large dataset. In contrast, data from the blogosphere is huge and much more complex [models of blogosphere described in (Gliwa et al. 2012)]. We can analyse not only the relations between authors, but also examine the emotions [sentiment analysis (Gliwa et al. 2012)] as well as topic modeling (Blei et al. 2003).

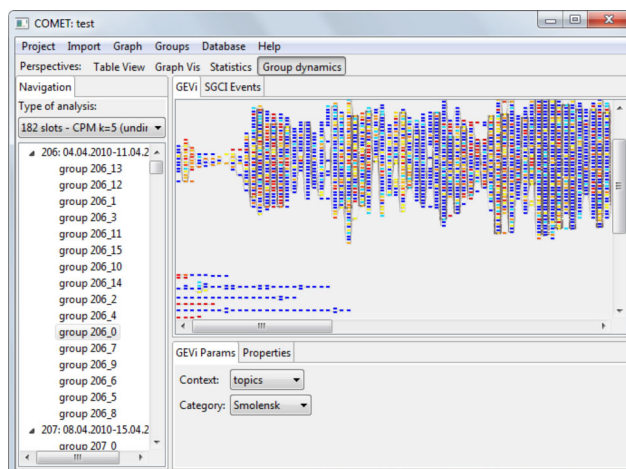


Fig. 9 Visualisation of contexts in GEVi

7.1 Dataset description

Tool capabilities will be presented based on data set contains data from the portal *salon24*⁶. 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 presented results were conducted on half of this dataset—from 4.04.2010 to 31.03.2012. The analyzed period was divided into time slots, each lasting 7 days and neighboring slots overlap each other by 4 days (numbered from 206 to 387). In the examined period there are 182 time slots. In each slot we used the comments model, introduced by us in (Gliwa et al. 2012)—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.

7.2 Group extraction and evolution

After separation of time slots we extracted the groups in each time slot. We used CPM method of community extraction (CPMd version from CFinder tool⁷) for $k = 5$.

Transitions between groups were assigned using our method SGCI described earlier. The threshold on modified Jaccard measure was set on level equals 0.5.

7.3 Sentiment calculation

The sentiment for posts and comments was calculated using a tool developed at the Luminis Research company⁸. Their method is based on searching words from analyzed text in a dictionary and calculating sentiment for found ones. More detailed description of this method we provided in (Gliwa et al. 2012).

The final value describing the overall sentiment is between -1 and 1 , but thresholds for negative, neutral and positive sentiment need adjusting. This can be done by analyzing some texts (part of texts earlier marked by algorithm) by human, manually assigning sentiment values (positive/negative/neutral) for them, next comparing these values with algorithm ones and finally setting appropriate thresholds.

In order to adjust thresholds for sentiment values, we analyzed about 150 random texts and based on this analysis we set the following thresholds: negative (<0), neutral ($0-0.3$), positive: (>0.3).

⁶ mainly focused towards politics, <http://www.salon24.pl>.

⁷ <http://www.cfinder.org/>.

⁸ <http://www.luminis-research.com>.

7.4 Topic modelling

Topics were extracted by LDA method (Blei et al. 2003) with using Mallet⁹ tool. After extraction there were 350 topics and then they were manually merged into some groups and labelled. Finally, we got 31 topics that were used further.

7.5 Integration sentiment and topic modelling with SNA

Every interaction between people (using comment model) is enhanced with information about sentiment and found topics in comments written by users. For given group all interactions between members of this group are taken into consideration. For each sentiment type (positive, negative, neutral) and each topic we counted interactions matching to these types and finally we assigned intensities for these types (these types are called *categories*; sentiment and topics we called *contexts*) as percent of interaction falling to given category in relation to all interaction within given group. For topics we reduced matching topics to these that intensity in given group is above 5 %.

7.6 Group sizes

As we can see on Fig. 10 most groups are small. Groups with size equal 5 outnumber others.

Using GEVi, we can observe the size for each group as it was demonstrated on Fig. 5. For instance, the group 311_7 has 5 members and size of group 312_4 equals 5.

7.7 Number of groups in timeslots

Figure 11 shows how number of groups with given size changes in time slots. We can observe that highest fluctuations in quantity have groups with size equal 5.

In GEVi the number of groups in each time slot can be easily noticed—the groups from the same time slot in the same hierarchy are positioned vertically one above the other.

7.8 Stability of groups in timeslots

In Fig. 12, mean stability between groups in slots is presented (e.g., stability in the slot 300 corresponds to stabilities between groups from the slot 300 and the slot 301). We can observe that stability has highest value around slot 210 and as we can see on Fig. 11 in that period there is very few groups.

⁹ <http://mallet.cs.umass.edu/>.

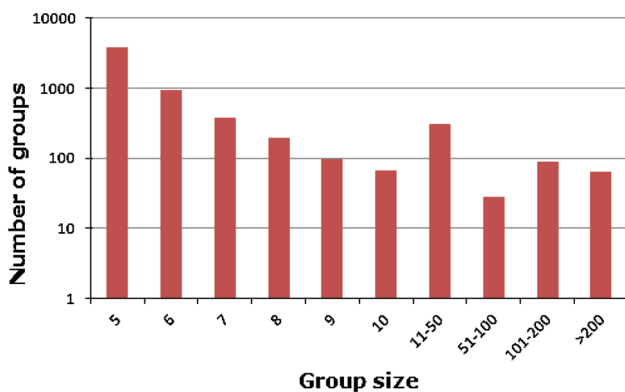


Fig. 10 Number of groups with given size

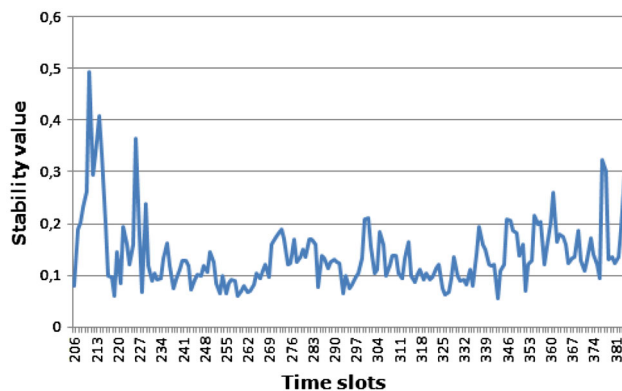


Fig. 12 Stability of groups in time slots (mean value)

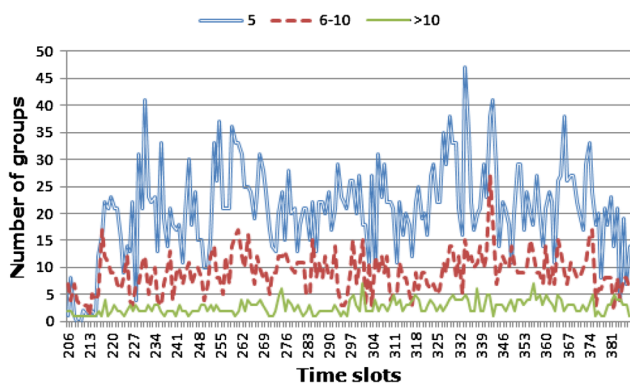


Fig. 11 Number of groups with given size in time slots

The stability of each transition between groups can be observed in GEVi when hovering mouse pointer over a certain chosen group—see Fig. 6, or indirectly: if in a given time slot there are more dashed transition arrows, the mean stability is expected to be less than in timeslots when there are mainly solid arrows.

7.9 Exchange of members of group in time

Some different hierarchies can be visualised in GEVi. The most interesting one is shown in Fig. 13, where the highlighted groups are the ones having in common at least one member with the first group in this hierarchy (group labelled as 206_4). The mentioned group (which is the biggest in its time slot) has 97 members and as we can notice, in each next time slot (every time slot has different vertical layer in visualization) there is at least one group that has any common members with that group (what is presented in Fig. 13). We can also observe that many groups overlap with this group (also in the same time slot).

This example shows how this tool can be used in analyzing, how long a given group can exist without complete exchange of initial members of group.

7.10 Common members between groups in the same time slot

Figure 14 presents summary of common members in group pairs from the same time slot. It seems that very few groups have common members with more than 2 groups—most pairs (from each possible in every time slot) have not common members.

We can also observe on Fig. 15 the distribution of common members between pair of groups in time. If we compare this figure with Fig. 11 then we can notice that peaks are in similar time slots on both charts. It suggests that increasing number of groups results in increasing of overlapping groups.

GEVi makes possible checking common elements for each selected group with the other ones. For instance, in Fig. 7 we can see that group 214_7 has 8 members and with group 215_0 has 5 common members, with 214_10 has 3 common member and there is no common members with group 214_16.

7.11 Overlapping groups in the same time slot

Figure 16 presents how much groups overlap other ones in the same time slot. We can notice that mostly groups overlap quite heavily. Most groups overlap with 4–6 other groups. It seems that there are very few groups completely isolated (without overlapping).

The presented tool enables possibility to check the group overlapping in the same time slot. Referring to Fig. 7, one can see that the group 214_7 overlaps with 2 other groups in the same time slot.

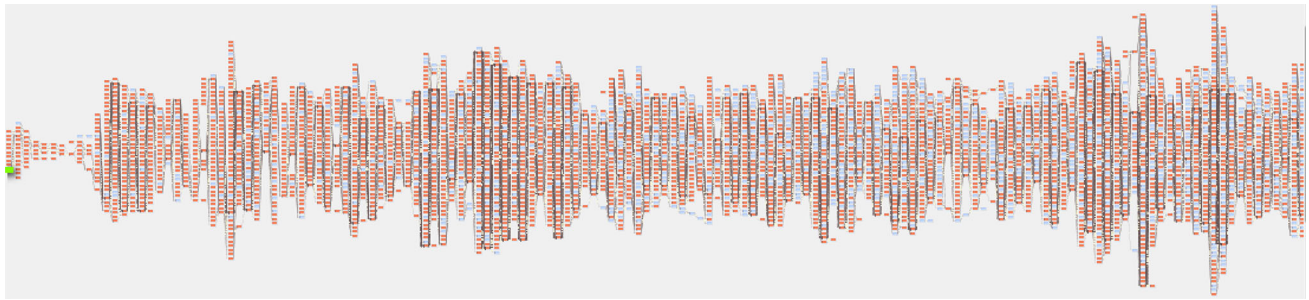


Fig. 13 Visualisation of groups that have common members with first group in hierarchy

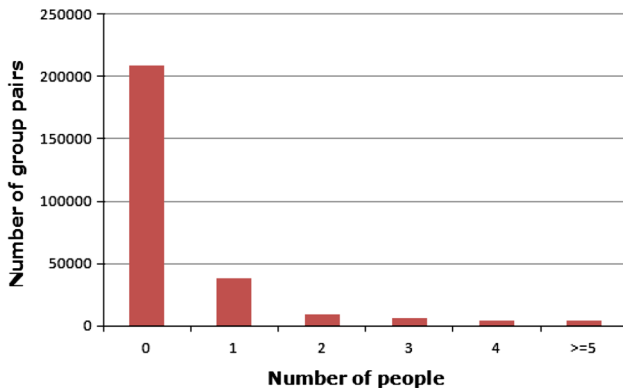


Fig. 14 Common members in group pairs from the same time slot

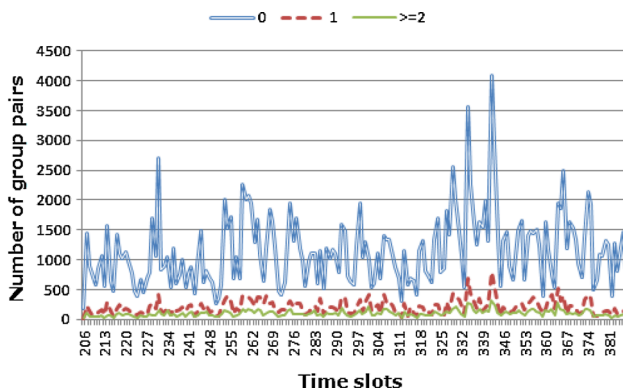


Fig. 15 Common members in group pairs in time slots

7.12 Topics in groups

In analysed period Polish blogosphere was highly influenced by one important event—Polish President airplane crash in Smolensk and some other events related with investigation of this catastrophe. Therefore, we identified some key events:

- Smolensk crash—slot 207 [10.04.2010]
- Initial MAK report—slot 217 [19.05.2010]
- Final MAK report—slot 275 [12.01.2011]
- Smolensk crash anniversary—slot 298 [10.04.2011]
- Smolensk Miller report—slot 326 [29.07.2011]

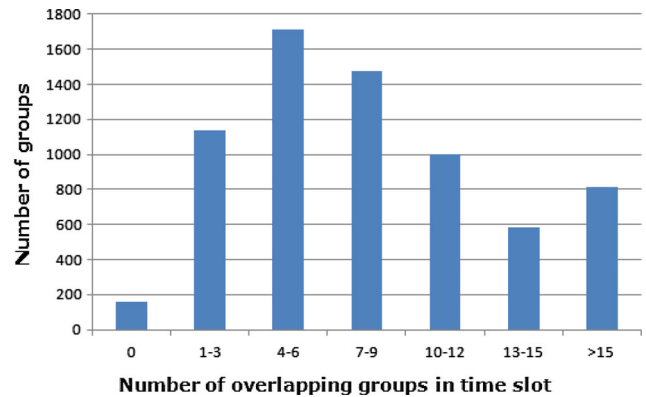


Fig. 16 Number of groups that overlaps with given number of other groups in the same time slot

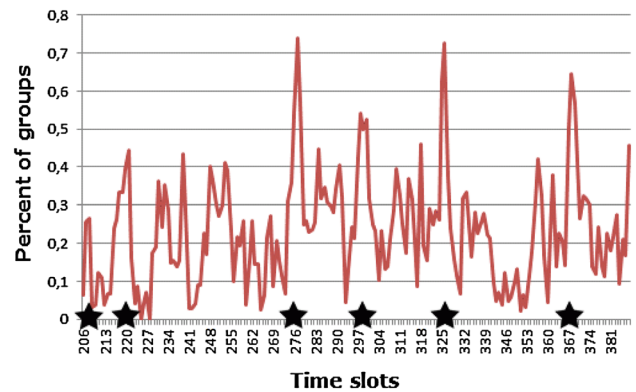


Fig. 17 Percent of groups in time slot that discuss about topic *Smolensk*

- Expertise of black boxes from crashed plane—slot 367 [16.01.2012]

Figure 17 shows percent of all groups in each time slot that discuss about topic of *Smolensk* (context: *topics*, category: *Smolensk*). Key events are marked on this figure by stars. We can see that the figure contains peaks around mentioned events.

Figure 18a and b (parts are overlapping) present groups with selected topic as *Smolensk*. As we can observe, this

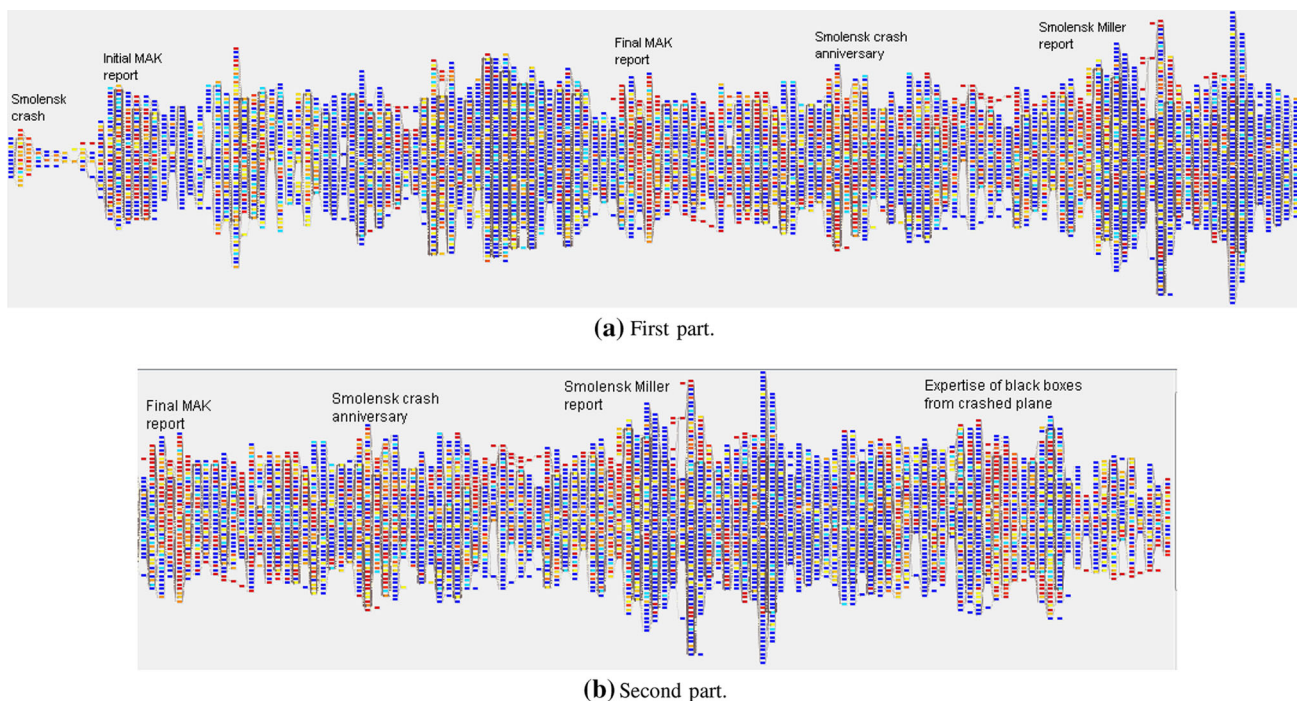


Fig. 18 Visualisation of intensity of topic *Smolensk* in groups (parts are overlapping)

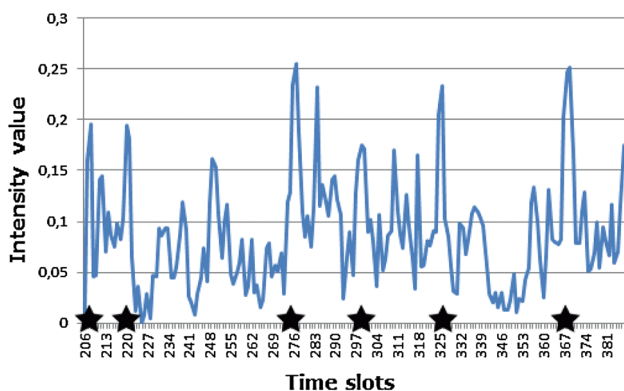


Fig. 19 Mean intensity of topic *Smolensk* in groups in time slots

topic is highly induced by events in real world related with *Smolensk* airplane crash. One can notice that near mentioned key events the topic *Smolensk* in groups has higher intensity (*red color* means the highest intensity).

Similar observation can be performed using Figure 19 which shows the mean value of intensity of topic *Smolensk* in time slots. We can discern that Figure 19 is very similar to Figure 17. It confirms that topic *Smolensk* has big impact on discussions carried out by people in groups.

Figure 20 demonstrate features of GEVi to analyse groups dynamics on different levels of details. We chose

event *Final MAK report* and in that place the view can be zoomed so we can look at it to see more details about group transitions. Green ellipses mean that these particular places are zoomed again and the result is presented on left and right side. Yellow ellipses mark biggest groups in their time slots. So we can observe that biggest group in slot 274 has 42 members and biggest group in slot 275 has 112 members. Groups are colored according to intensity of topic *Smolensk* and we can notice that biggest group in slot 274 (group 274_10) has low intensity but biggest group in slot 275 (group 275_7) has very high. So between these 2 slots the group 274_10 highly increased in size and changed a lot topics of discussions. On the left there are some groups where this topic is very intense (one of them has size 28) - they are merging with group 274_10 and during these events this group grows and this topic is more popular among members of this group.

Figure 21 shows selected topic *Elections*. We can also observe some events increasing popularity of this topic in time:

- Presidential elections—20.06.2010
- Local government elections—21.11.2010
- Parliamentary elections—9.10.2011

However, these events don't have such big impact as it was shown for topic *Smolensk*.

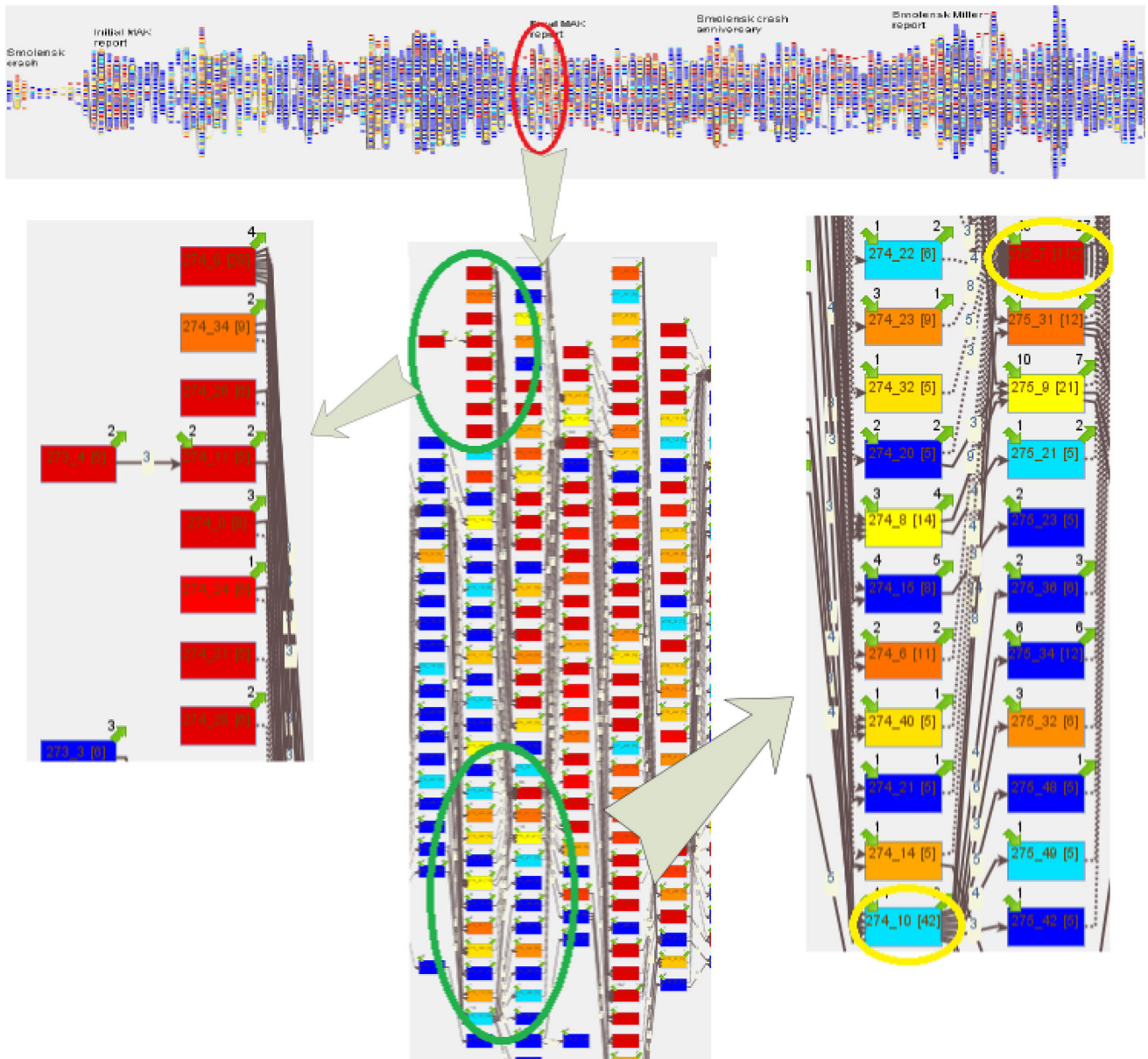


Fig. 20 Visualisation of different zooming levels to analyze chosen key event in topic *Smolensk*

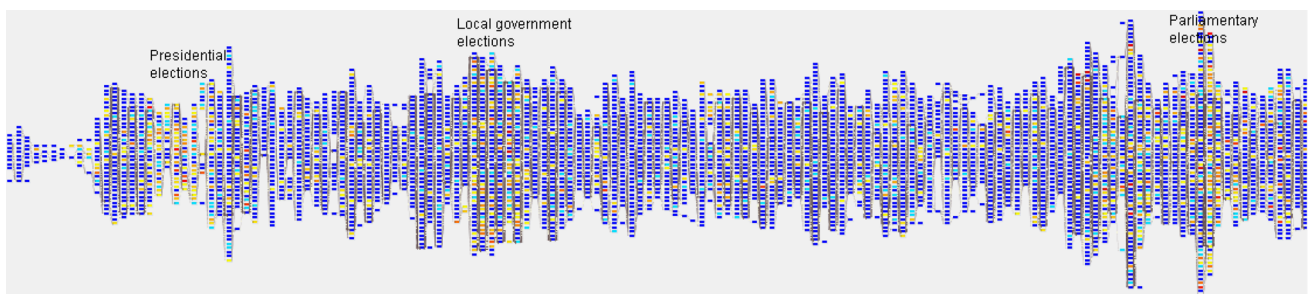


Fig. 21 Visualisation of intensity of topic *Elections* in groups

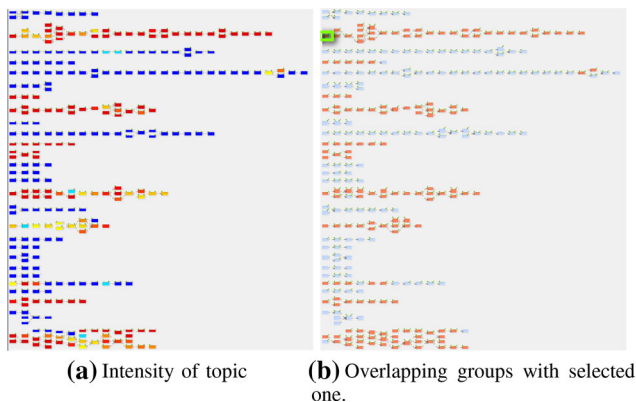


Fig. 22 Different views for groups discussing about *Science and education*

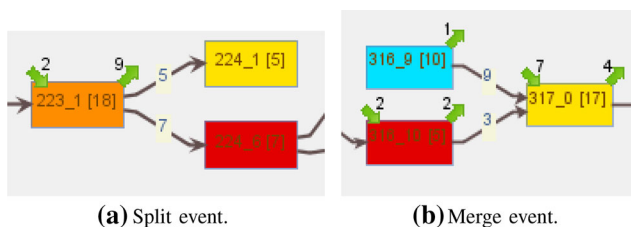


Fig. 23 Selected events for groups with visualisation of intensity of topic *Science and education*

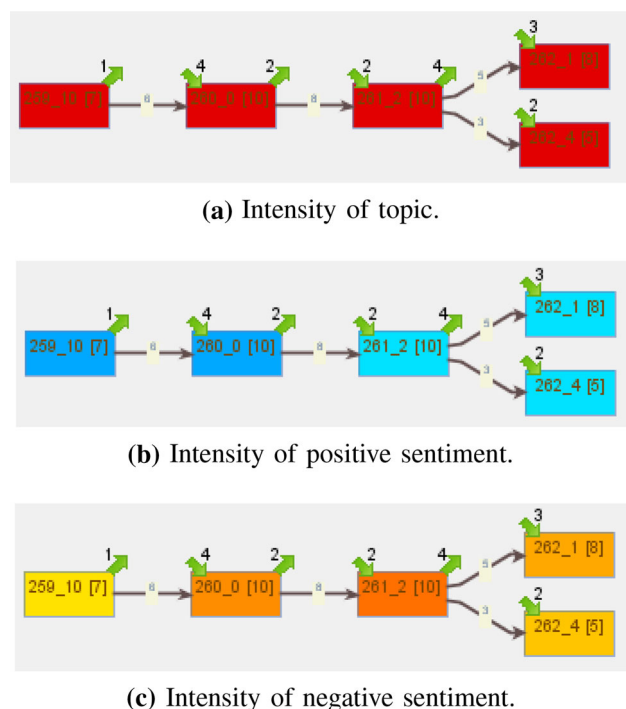


Fig. 24 Different views for groups discussing about topic *Smolensk*

Table 1 Comparison of visualisation tools

Feature	Tool from Reda et al.	ViENA	SnailVis	GEVi (our)
Visualisation of groups in time	+	+	+	+
Visualisation of group structure	-	+	-	+/- (only list of members)
Visualisation of group lifecycle events	-	-	-	+
Visualisation of migration on single user level	+	-	-	-
Visualisation of additional aspects of groups	-	-	-	+
Support for overlapping groups	-	-	-	+

Another interesting example is topic *Science and education*, shown on Fig. 22a. There are visualised some hierarchies of groups with this topic and we can observe that this topic is very stable in time (if first group in hierarchy has intense this topic, then in next slots this topic is also highly discussed by members of groups). On Fig. 22a and b we can compare groups that have high intensity of this topic and overlapping groups with one of group discussing this subject. There is huge correlation between these figures. It means that topic *Science and education* is mostly connected with particular people and if they are in many groups then this topic is also discussed there.

Using GEVi we can also make some other interesting observations on this topic. For example, when group talking about this topics splits, usually in one of the resulting group this topic is much less intensive (Fig. 23a) but after merging some groups the intensity of created group is lower than the most intense from merging groups.

7.13 Sentiment in groups

Figure 24a presents small group talking mainly about *Smolensk crash*. One can see that in this group there is more negative sentiments (Fig. 24c) than positive ones (Fig. 24b). It means that such controversial topics like *Smolensk* arouse many emotions (and also many negative ones). We compared 2 selected topics: *Smolensk* (as an example of controversial topic) and *Recreation and hobby*, which is presented in Table 1. We can see that for *Smolensk* negative part is larger than positive one, but for *Recreation and hobby* the situation is reversed.

In Fig. 25 there is example of group with high part of positive sentiment—as we can see the topics are rather non-controversial.

Table 2 Comparison of sentiments (mean/standard deviation) for 2 selected topics: *Smolensk* (controversial) and *Recreation and hobby* (non-controversial) [mean/stdDev]

Sentiment	Smolensk	Recreation and hobby
Positive	0.155/0.064	0.222/0.062
Negative	0.204/0.07	0.109/0.046
Neutral	0.641/0.072	0.669/0.061

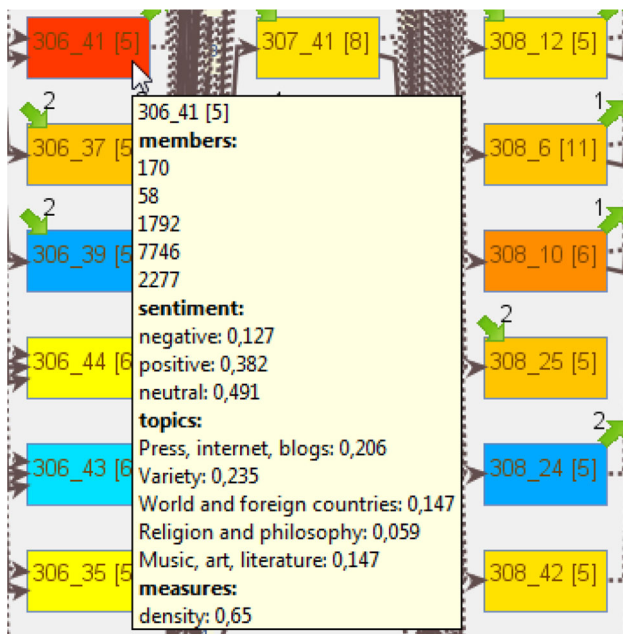


Fig. 25 Example of group with high part of positive sentiment

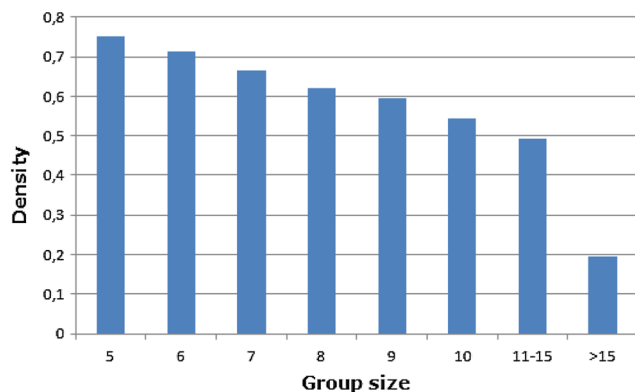


Fig. 26 Mean density for groups with given size

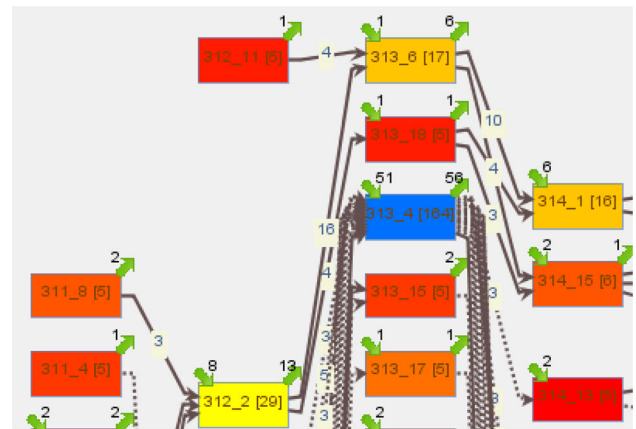


Fig. 27 Visualisation of intensity of groups density

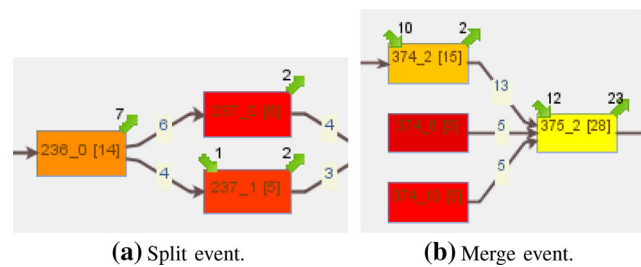


Fig. 28 Selected events for groups with visualisation of intensity of density

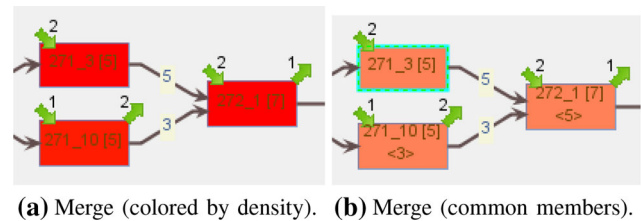


Fig. 29 Visualisation of merge for groups with high overlapping between merging groups

7.14 Groups density

Figure 26 presents relation between group size and their mean density. One can see that density is decreasing when the group size is increasing.

Similar observation we can perform using GEVi tool on local level. Figure 27 presents groups coloured according to their density. We can observe that large groups (with more than 100 members) have significantly lower density than small groups.

After splitting mostly the groups have more densities than splitting group (Fig. 28a) and with merge the resulting group has usually lower density than both merging groups (Fig. 28b). But if 2 groups have a lot of common members and these groups merging, the resulting group have very

similar density to them (Fig. 29a,b)—groups 271_3 and 271_10 have 5 members, but 3 of them are common.

8 Conclusion and future directions

In this paper GEVi's features were described. GEVi can be used standalone or as a part of any tool (currently is integrated with tool for complex network analysis COMET). The tool allows to analyse dynamics of group with taking into consideration context of groups. It enables also analyse group dynamics on different levels of details (analysis top-down) and can be useful tool to analysis of impact of key events on network dynamics. GEVi provides analysis of different aspects of groups and their influence on groups dynamics. It enables better understand groups and their dynamics.

In the future we plan to add possibilities of detecting new evolution events, enable visualisation of groups dynamics on the level single person [with taking into account roles played by users in different groups, especially roles defined by us for blogosphere (Gliwa et al. 2013)] and to employ other real-world data to tune-up the proposed network analysis tool.

Acknowledgments This publication is based on work supported by Research project No. O ROB 0008 01 "Advanced IT techniques supporting data processing in criminal analysis", funded by the Polish National Centre for Research and Development. The authors thank P. Maciołek who provided and allowed the use of the algorithm and tools for analysis of sentiment of texts in Polish language. The authors also thank S. Podgórski who calculated topics for salon24 dataset and prepared them to use.

References

- Agarwal N, Liu H (2009) Modeling and data mining in blogosphere. Synthesis lectures on data mining and knowledge discovery. Morgan and Claypool Publishers, San Rafael
- Asur S, Parthasarathy S, Ucar D (2009) An event-based framework for characterizing the evolutionary behavior of interaction graphs. *ACM Trans Knowl Discov Data* 3(4):16
- Bastert O, Matuszewski C (2001) Layered drawings of digraphs. In: Kaufmann M, Wagner D (eds) *Drawing graphs*. Springer, Berlin, pp 87–120
- Beiro MG, Busch JR, Alvarez-Hamelin JI (2010) Visualizing communities in dynamic networks. In: *LAWDN-Latin-American workshop on dynamic networks*. Buenos Aires, Argentina
- Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *J Mach Learn Res* 3:993–1022
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008(10)
- Borgatti SP, Everett MG (1993) Two algorithms for computing regular equivalence. *Soc Netw* 15(4):361–376
- Bródka P, Saganowski S, Kazienko P (2013) Ged: the method for group evolution discovery in social networks. *Soc Netw Anal Min* 3(1):1–14
- Evans T, Lambiotte R (2009) Line graphs, link partitions and overlapping communities. *Phys Rev E* 80(1 Pt 2):016105
- Federico P, Pfeffer J, Aigner W, Miksch S, Zenk L (2012) Visual analysis of dynamic networks using change centrality. In: *2012 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, pp 179–183
- Fortunato S (2010) Community detection in graphs. *Phys Rep* 486(3):75–174
- Gansner ER, Koutsofios E, North SC, Vo K-P (1993) A technique for drawing directed graphs. *IEEE Trans Softw Eng* 19(3):214–230
- Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc Natl Acad Sci* 99(12):7821–7826
- Gliwa B, Kozlak J, Zygmunt A, Cetnarowicz K (2012) Models of social groups in blogosphere based on information about comment addressees and sentiments. In: *4th international conference on social informatics, SocInfo, Lausanne, Switzerland*. Lecture Notes in Computer Science, vol 7710. Springer, Berlin, pp 475–488
- Gliwa B, Saganowski S, Zygmunt A, Bródka P, Kazienko P, Kozlak J (2012) Identification of group changes in blogosphere. In: *The 2012 international conference on advances in social network analysis and mining, ASONAM 2012*. IEEE Computer Society
- Gliwa B, Saganowski S, Zygmunt A, Bródka P, Kazienko P, Kozlak J (2012) Identification of group changes in blogosphere. In: *IEEE/ACM international conference on advances in social networks analysis and mining, ASONAM 2012, Istanbul, Turkey*
- Gliwa B, Zygmunt A, Byrski A (2012) Graphical analysis of social group dynamics. In: *CASoN*. IEEE, pp 41–46
- Gliwa B, Zygmunt A, Kozlak J (2013) Analysis of roles and groups in blogosphere. In: *8th international conference on computer recognition systems, CORES 2013, Milkow, Poland, 27–29 May 2013*. *Advances in intelligent and soft computing*, vol 226. Springer, Berlin, pp 299–308
- Greene D, Doyle D, Cunningham P (2010) Tracking the evolution of communities in dynamic social networks. In: *Proceedings of the ASONAM '10*. Washington, DC, USA: IEEE Computer Society
- Hanneman RA, Riddle M (2005) *Introduction to social network methods*. University of California, Riverside
- Jung JJ (2011) Boosting social collaborations based on contextual synchronization: an empirical study. *Expert Syst* 38(5):4809–4815
- Macskassy S (2011) Contextual linking behavior of bloggers: leveraging text mining to enable topic-based analysis. *Soc Netw Anal Min* 1(4):355–375
- Mostafa M (2013) An emotional polarity analysis of consumers airline service tweets. *Soc Net Anal Min* 3(3):635–649
- Palla G, Barabasi A-L, Vicsek T (2007) Quantifying social group evolution. *Nature* 446(7136):664–667
- Palla G, Derenyi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435:814–818
- Porter MA, Onnela J-P, Mucha PJ (2009) Communities in networks. *Not Am Math Soc* 56(9):1082–1097
- Reda K, Tantipathananandh C, Johnson AE, Leigh J, Berger-Wolf TY (2011) Visualizing the evolution of community structures in dynamic social networks. *Comput Graph Forum* 30(3):1061–1070
- Spiliopoulou M (2011) Evolution in social networks: a survey. In: CC Aggarwal (ed) *Social network data analytics*. Springer, Berlin
- Wasserman S, Faust K (1994) *Social network analysis: methods and application*. Cambridge University Press, London
- Zygmunt A, Bródka P, Kazienko P, Kozlak J (2012) Key person analysis in social communities within the blogosphere. *J UCS* 18(4):577–597