Discovery and Exploitation of Knowledge in Collaboration Social Networks

Gabriel Tutoky*
Department of cybernetics and artificial intelligence
Faculty of Electrical Engineering and Informatics
Technical University of Košice in Košice
Letná 9, 042 00 Košice, Slovakia
gabriel.tutoky@tuke.sk

Abstract

This article deals with the discovery and exploitation of knowledge stored in the collaboration social networks (SN). We analyse available techniques for projection of Affiliation Networks (AN). In this process three partial tasks can be identified - projection of collaborations, projection of parallel and projection of sequential events, all influencing on the information reduction in the projected networks. For projection process we designed the method for the generation of weighted relations among participants of single event so that the information about number of event participants and event duration is considered. In the projection task of parallel events we analyze the influence of multiplex relations and in the projection task of sequential events we analyzed and designed a modification of the relation aging method. This results from the assumption that relations among actors in the network are not constant in time, but in the case it does not come to their regular updating, their strength decreases. Classical as well as newly proposed approaches to weighting of ties in collaboration networks are experimentally evaluated. We performed three experiments with real data set and as a reference we used data gathered from peoples' opinions expressed in targeted inquiries.

Categories and Subject Descriptors

E.1 [**Data Structures**]: Graphs and networks; G.2.2 [**Discrete mathematics**]: Graph Theory—*Graph algorithm*; H.2.8 [**Database Management**]: Database Applications—*Data mining*; I.2.1 [**Artificial Intelligence**]: Applications and Expert Systems—*Industrial automation, Office automation*; I.2.m [**Artificial Intelligence**]: Miscellaneous

Keywords

Social Network Analysis, Network Projection, Collaboration Networks, Data-mining, Relation ageing

1. Introduction

In recent years, many "social networks" have been analyzed like various Internet communities, email networks, peer to peer networks, telephone call graphs or traffic train routes [11]. All of these networks are interesting in some of their specific aspects and they provide a notable data source for network analysis. There are usually large-scale networks with thousands of nodes and edges. Analysis of these networks, usually based on global properties, can lead to interesting and helpful results. Nevertheless, there exist many different situations in the network analysis where data used for analysis of these networks did not carry sufficient information, e.g. temporal information is often neglected in these analyses. In this paper we propose a new approach how to model and analyze one particular type of social networks - affiliation networks (AN) making use of more strands of additional information, including the temporal one.

An AN is a network of actors connected by common memberships in groups/events such as clubs, teams, organizations or some common activities. AN are special type of two-mode social networks [16] where one mode is a set of actors, and the second mode is a set of events which are affiliated to the actors. A tie between an actor and an event is created, if this actor participates on particular event. Affiliation networks describe collections of actors rather than simply ties between pairs of actors. Based on such an affiliation network we are able to derive connections among members of one of the modes based on linkages established through the second mode [16].

Affiliation networks were studied in past, e.g. studying attendance of women in social events [5], movies and their actors [17] or co-authorship network of scientists and their papers [11, 10]. Whereas in the first two examples, the authors used unweighted representations of the networks, in the last work, the author introduced interesting approach for building of collaboration network with weighted ties between authors of the same paper. The weight of the tie between collaborating authors of a single paper is derived from a count of the papers coauthors and final weight of two collaborating authors is a sum of weights over all the papers where authors collaborated. This approach allows e.g. finding the "most connected" scientists in the whole collaboration network.

*Recommended by thesis supervisor: Assoc. Prof. Ján Paralič. Defended at Faculty of Electrical Engineering and Informatics, Technical University of Košice, Košice on September 21, 2011.

Tutoky, G. Discovery and Exploitation of Knowledge in Collaboration Social Networks. Information Sciences and Technologies Bulletin of the ACM Slovakia, Vol. 3, No. 4 (2011) 28-36

In our work we build collaboration network of teenagers (the necessary background details are described in Appendix A) based on their participations on educative pedagogic workshops. Created network we next project onto one-mode network which is more suitable network type e.g. for evaluating of the "most important" persons in desired time stamp or e.g. for further exploitation of information stored in the projected (or modeled) network. In the network projection process we identify three sub-tasks and for each one we propose new projection approaches or some modifications of existing aproaches and subsequently we evaluate them.

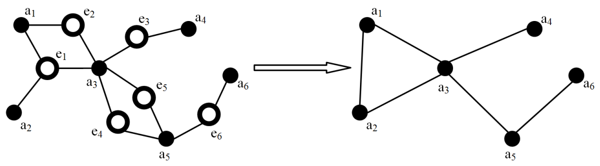

Figure 1: Projection of two-mode network (• - actors, ○ - events) onto one-mode network from actors' view.

2. Projection of Affiliation Networks

Affiliation networks (two-mode representation) are most often projected onto one-mode networks for their further analysis [2]. Projection allows analysis of network from one of the two possible perspectives - *actors' view* or *events' view*. This property is also called as *duality* of two-mode networks. In actors' view two actors are connected if they participated together in at least one event, so they *cooperated* together. In events' view two events are connected if at least one actor participated on both events. It is called *overlapping of events* [16].

Consider an affiliation network represented by bipartite graph $G = (A, E, T)$ [1]. The A-projection (actors view) of G is the graph $G_T = (A, T_A)$ in which two nodes (of A) are linked together if they have at least one neighbor in common (in E) in G: $T_A = \{(a_i, a_j), \exists x \in E : (a_i, x) \in T \wedge (a_j, x) \in T\}$, where $a_i, a_j \in A$. The E-projection is defined dually [8]. See Figure 1 for an example.

2.1 Base projection methods

Usually, weights in both, affiliation (two-mode) and also in projected (one-mode) networks have binary values. The ties in the networks exist or not [3]. In the network projection process of two-mode networks onto one-mode networks we can use different measures for weight calculation, e.g. the ones summarized by Opsahl in [12]:

- Weight is determined as a count of parrticipations (co-occurrences) - e.g. count of the events were two actors participated together, formalized expression is:

$$w_{ij} = \sum_{e \in E} 1 \tag{1}$$

  where w_{ij} is the weight between actors i and j (nodes of the first mode), and e are events (nodes of the second mode) where i and j participated together.

- Newman in [11, 10] proposed extended determination of weights while working with scientific collaboration networks. He supposes that strength of social bonds between collaborators is higher with lower number of collaborators on a paper and vice versa social bonds are lower with many collaborators on a paper. He proposed formula (see formula 2) for defining the weights among collaborators where N_e is the count of collaborators on paper (event) e.

$$w_{ij} = \sum_{e \in E} \frac{1}{N_e - 1} \tag{2}$$

- Till now we considered only binary two-mode networks and their projection to weighted one-mode networks. However, there exist also weighted two-mode networks, such as networks of online forums (weight is determined as count of posts or posted characters) or collaboration network described above and also in [14, 15]. So, both just presented measures for weight definition could be extended for weighted two-mode networks as follows:

$$w_{ij} = \sum_{e \in E} w_{j,e} \tag{3}$$

  where $w_{j,e}$ is the weight of j^{th} actor to e^{th} event where i and j participated together. This method differentiates how two particular actors interact with the common event, and projects this information onto a directed weighted one-mode network [12].

- In a similar way, the Newman's method can be extended for projection of weighted two-mode networks. The weights are calculated by the following formula:

$$w_{ij} = \sum_{e \in E} \frac{w_{j,e}}{N_e - 1} \tag{4}$$

  This formula leads to a directed one-mode network in which the out-strength of a node is equal to the sum of weights attached to the ties in the two-mode network that originated in that node [12].

2.2 Sub-processes in network projection

Projection of two-mode networks is often considered as one single process, however by analysis of this process with considering temporal attributes we identified three sub-processes which are notable for final projection results:

- *Projection of collaborations* - identifies participants of a single event and assigns weight of the tie between them. Weight is determined by event properties, e.g. by number of event participants or by event duration.
- *Projection of parallel events* - in several types of networks, actors can participate on parallel events, e.g. two directors are members in two board of directors together in the same time. Strength of the tie between these actors can be computed by several ways, e.g. as maximum or average value of all identified ties or we can apply principle of multiplex and layered networks [7].
- *Projection of sequential events* - more often in the networks we can identify sequential events which are ordered by their temporal attributes. From this we are able to compute frequency of collaborations or time spent between two events for selected pair of actors.

In the next sections we particularly describe our extensions for projection of two-mode networks onto one-mode networks. Projection of two-mode networks has strong impact on analysis of collaboration networks. It is important step for creation of the most suitable network model by one-mode projection method.

3. Projection of collaborations

At first, we propose new, more general weighting of the ties created among event participants as Newman's weighting method. The reason is that Newman's weighting method results in fast decreasing value with just a small increase of event participants (more than two). This can be good in some cases, but not in general for any collaboration network. We suggest using one of the two proposed types of curves - exponential (section 3.1) or sigmoid (section 3.2).

3.1 Exponential weighting

The weights are also decreasing with increasing number of event participants like in Newman method, but decay parameter α can be adjusted with respect to particular type of collaboration network (and in such a way influence the shape of the exponential curve). Formally:

$$w_{ij}^{p} = \alpha_e^{2-N_e} \quad (5)$$

where parameter α depends on collaboration type and it should be estimated by a domain expert with the following formula:

$$\alpha_e = \sqrt[\beta-2]{2} \quad (6)$$

where β is the number of participants (event size) when weight of collaboration ties decreases by 50%. This formula enables easier set up of an optimal value of the parameter α for particular type of collaboration network. For example in scientific collaboration network, the authors of an article (participants of an event) usually cooperate in smaller groups than e.g. actors in the DAKCSN network described in Appendix A. It causes that different number of participants establishes links with 50% of maximum weight. Number 2 used in the index of radical in equation (6) represents an "ideal" number of event participants when the strongest ties (with value 1) are created among event participants (this is analogical to the Newman's method).

3.2 Sigmoid weighting

This method assigns collaboration weights with small difference to the previous exponential method. In this case we proposed slower decrease of the collaboration weights with small increase of event size. With additional increase of event size, the progress of collaboration weights is almost linear. It is expressed by the following formula:

$$w_{ij}^{p} = sigm\,(N_e) = \frac{1}{(1+\varepsilon) + (1+\varepsilon)\alpha_s^{(\beta-x)}} + 1 \quad (7)$$

where α_s is decay factor of sigmoid curve and β is number of participants when collaboration weight decreases by 50% (point where sigmoid changes its character from concave to convex). Decay factor α_s should be computed by:

$$\alpha_s = \sqrt[\beta-\gamma]{\frac{(1-0.75)^{-1} - (1+\varepsilon)}{1+\varepsilon}} \quad (8)$$

where γ is the number of participants where their collaboration weight is 75% of maximal value.

For each vale of β and γ factors equation (7) does not guarantee maximum collaboration weight (of value 1) for event with two participants and there is suitable to normalize weight by formula:

$$w_{ij} = \frac{sigm(N_e)}{sigm(2)} \quad (9)$$

3.3 Temporal weighting

This weighting type consider duration time of events when participants cooperates together. The weight is computed by

$$w_{ij}^{t} = \kappa_t T_e \quad (10)$$

where T_e is the event duration time and κ_t is time normalization factor $\kappa_t = \frac{1}{T_{opt}}$ and T_{opt} is the optimal time of event duration with two participants when they establishe relations with maximum strength. Presented method should be combined with one of the previous ones by formula:

$$w_{ij} = w_{ij}^{p}.w_{ij}^{t} \quad (11)$$

where w_{ij} is the final collaboration weight.

3.4 Results

3.4.1 Evaluation methods

Collaborations modeled by presented methods we evaluated on real data set - DAKCSN-Real (described in Appendix A.3) in three ways:

- *Standard classification* - collaboration weights are classified into one of the interval which represents collaboration strength and next we compare modeled collaborations with real data (derived from inquiries). Classification is evaluated by standard metrics.
- *Classification with tolerance* - we utilize ordinary sorted intervals of collaboration strength. We modified evaluation of classification so that we considered the ties classified into neighbor intervals (by comparison to real data) as properly classified.
- *Classification distance* - for each collaboration tie we compute distance between modeled and real weight (middle point of interval where respondents assign the tie) and next we compute mean absolute deviation of all ties.

3.4.2 Comparison of weighting methods

Our results for weighting curves proposed in sections 3.1 and 3.2 show sensitivity of classification results to the factor β. The best results were reached for values $\beta = 6$ and $\beta = 9$ for exponential and $\beta = 9$, $\gamma = 6$ for sigmoid weighting curve (see Figure 2).

Results of mean absolute deviation (Figure 2 in window) show that minimal average classification distance is 0.276. At average this value falls into neighbor classification interval and so we evaluated also classification with tolerance (see above). It resulted to 61.2% of classification electiveness instead of 27.9% to direct classification. For comparison of all metrics see Table 1.

Table 1: Comparison of classification effectiveness results of particular metrics with and without considering duration time of events (T_u).

Metrics	Micro F1		Micro F1 with tolerance	
	no T_u	T_u	no T_u	T_u
Binary	6,23%	6,23%	21,72%	21,72%
Newman	22,40%	24,88%	51,40%	56,23%
Exponential (β=9)	**27,96%**	**26,06%**	**61,20%**	**57,72%**
Exponential (β=11)	24,85%	22,95%	56,84%	54,42%
Sigmoid (β=7, γ=3)	18,57%	16,86%	44,84%	42,72%
Sigmoid (β=9, γ=3)	22,12%	19,83%	49,02%	46,87%
Sigmoid (β=11, γ=6)	23,28%	20,37%	50,13%	47,75%

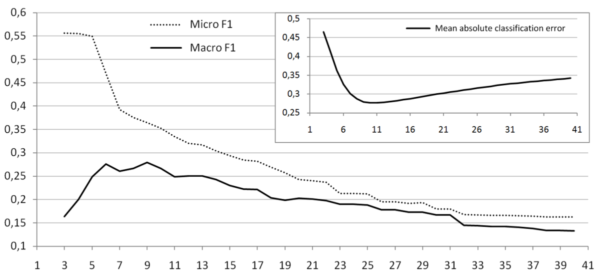

Figure 2: Dependency of classification effectiveness in micro and macro averaging to the factor β. In window: Dependency of mean absolute classification error to the factor β

4. Time based projection of events

Various collaboration networks contain time series data - usually time of the events are known. Considering this we recognize two main event types - *parallel* and *sequential* events. Actors of common parallel events cooperate together in multiple relations and their final cooperation is given by aggregating function of multiple relations (section 4.1). For sequential events it is reasonable to assume that weight of the ties created between participants of a common event will decrease over time. So, we propose time dependent weights in our representation of one-mode projected affiliation network - a kind of aging of the ties (section 4.2). This should be considered as similar approach to the one presented in [6, 18] where authors considered aging of the nodes in the context of citation networks. They describe node's age as influence to the probability of connecting current node to new nodes in the network.

4.1 Projection of parallel events

Projection of parallel events is similar to the principle of multiplex and layered social networks. It is based on the observation that in real world people cooperate in multiple roles (e.g. as father, employee, member of veteran club, etc.) where they cooperate with others and establish new social relations. One actor in such a way can establish various types of relations [7].

Aggregation of multiplex relations does not need to be simple process as depicted on Figure 3 a). In many cases we recognize relations of different influence (relations of family member are morally more valuable than business relations). So that each type of relations has assigned weight representing its strength or worth. Final weight

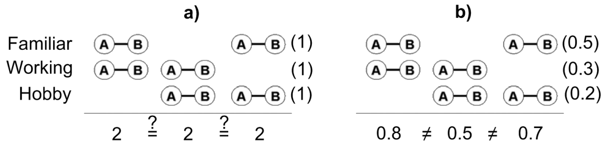

Figure 3: Aggregation of multiplex relations in layered networks a) - unweighted; b) - weighted relations[7].

aggregated from various relations depends on its weights (see Figure 3 b)), formally:

$$w_{ij} = \sum_{e \in E_p} w_{ij}^e . r^e \tag{12}$$

where E_p is a set of parallel events, w_{ij}^e is collaboration weight of particular event (e.g. assigned by number of event participants) and r^e is the weight of current relation type from event e.

4.2 Projection of sequential events

This projection of sequential events is based on relation weight ageing. Proposed method is based on assumption that past collaborations among network members are less important than lately created collaborations. These past collaborations after passing sufficient long time have no more influence in the present and they are next removed from the network[1] - old ties (without refreshing) among collaborators are than "forgotten" in such a way.

4.2.1 New opportunities and resembling work

From the social network analysis point of view our proposal of aging of the edges can lead to new opportunities in network analysis:

- *Tracking collaborations over the time* - i.e. tracking of collaboration strength with passing time among selected actors of the network. This should provide detailed information describing evolution of cooperation among desired actors.
- *Creation of network snapshots in given time* - it allows us to obtain actual network state in desired time and consequently to analyze e.g. strongest collaborations in the network. It can lead to different

[1]If the value decreases under an threshold value ε (some small number, e.g. 0.1 or 0.01)

results of network analysis because we do not consider older collaborations so important like last created. In collaboration network we are able to "view" still actual and (by our confidence) important collaborations among network members.

Resembling work is presented in [13] where authors simulate evolution of communication network for analysis of their active members. After creation of the relation (after sending of private message) they create the tie with value 1 which by passing time is decreasing exponentially by formula:

$$w_{ij}(t+\delta) = \left\{ \begin{array}{ll} w_{ij}(t)e^{-\theta\Delta t}, & \text{if } w_{ij}(t)e^{-\theta\Delta t} > \varepsilon \\ 0, & \text{otherwise} \end{array} \right\} \quad (13)$$

where $w_{ij}(t)$ is weight of the collaboration in time t and $w_{ij}(t+\delta)$ is collaboration weight after time δ. Value ε is threshold value of minimal collaboration weight. Factor θ is called *ageing factor* which designates "ageing speed". It is described by $\theta = \frac{ln2}{t_{1/2}}$, where $t_{1/2}$ is the time when weight of the ties decreases to 50% by aging process.

4.2.2 Network evolution

Evolution process (see Algorithm 1) is depicted on Figure 4 where $e_1, e_2, ..., e_m$ are events where actors i and j participated together. Aging of the ties allows terminate sporadical and nonsignificant relations (e.g. if two actors participated together rarely). Relations without periodical repetitions are removed from the network in passing time. Vice versa, in case of periodically repeated relations (we assume that these relations are significant), they are "highlighted" in the network so that their weight is increased (see events e_1 till e_3 on Figure 4).

Algorithm 1 Network evolution

Input: Events $e_1, e_2, ..., e_m \in E$ descending ordered by start time.
Output: Evolution of the network ties.

1. Apply projection of collaboration on event e_1 and create set of collaboration ties T.
2. For each $e_k \in E, k = 1, 2, 3, ..., m-1$
   (a) Apply projection of collaboration on e_{k+1} and create set of collaboration ties T_{k+1}.
   (b) If e_k and e_{k+1} are parallel events, i.e. $t_k = t_{k+1}$:
      i. Create $T = T \cup T_{k+1}$ so that apply projection of parallel events on overlapping ties.
   (c) If e_k and e_{k+1} are sequent events, i.e. $t_k < t_{k+1}$:
      i. Compute distance time δ_k between events e_k and e_{k+1}.
      ii. Apply ageing of the ties for each tie in T and set their value for time t_{k+1}.
      iii. Create $T = T \cup T_{k+1}$ so that compute sum of overlapping ties.

4.2.3 Proposed modifications

In contrast of aging process proposed in [13] where authors used constant ageing speed, we proposed some modifications for customizing ageing speed depending on strength or frequency. In first case we consider that not all relations ageing by the same speed, but it dependes on current relation weight. If weight is high, actors cooperate strongly and there are higher assumptions for their strong relation. In second case we consider dependency of relation weight on frequency of cooperation, so if two actors cooperate frequently even though with weak strength finally their relation will be strong. We consider so two properties which affect the ageing factor - *relation strength* and *previous collaborations*.

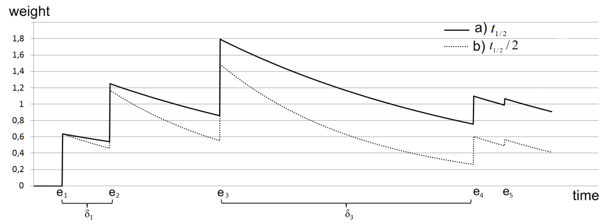

Figure 4: Evolution of the single tie between actors i and j for two various ageing factors.

Relation strength. In this case we apply proposed rule that ageing speed is slower for stronger collaboration relations. We will consider various ageing factor for relations of various strength created by projection of events. Ageing factor is based on relation strength (weight) by following formula:

$$\theta = \frac{ln2}{t_{1/2} w_{ij}(t)} \quad (14)$$

where $w_{ij}(t)$ is relation weight assigned from the last occurrence of cooperation between actors i and j.

Previous collaborations. In this second case we propose monitoring of previous collaborations from which we are able determine count, frequency and time between considered events where two actors i and j cooperated together. Ageing factor is so computed by formula:

$$\theta = \sum_{e \in E_{past}} \frac{r_\omega(\omega - \delta_e)}{\omega} \quad (15)$$

where E_{past} is the set of monitored events which were performed in previous period ω. Variable δ_e is passed time from the last performed event e and parameter r_ω is weight representing influence of previous relations to the ageing factor θ.

4.3 Results

In the first experiment we focus on evaluation of projection of parallel events. We assigned weights to all particular types of relations and we aggregate these multiplex relations. Our second experiment examines dependency of ageing factor to final model of collaboration network and in third experiment we observe influence of proposed methods for customization of ageing factor. Evaluation of the network projection (network model) we performed in ways described in section 3.4.1.

4.3.1 Parallel events

For our experiment we used the best metric - exponential (by the results presented in Table 1) and we monitored dependency of classification effectiveness to the factor β (see Figure 5). Results show (in comparison to same

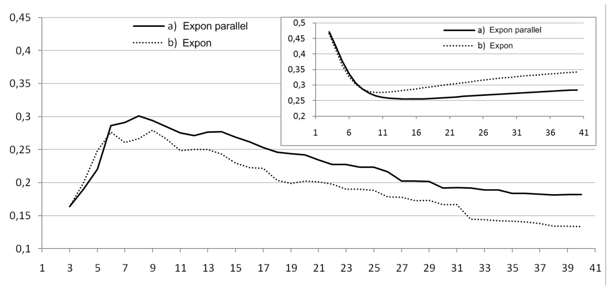

Figure 5: Influence of parameter β to the classification effectiveness in macro averaging for a) exponential metric with multiplexing; b) base exponential metric. In window: Influence of parameter β to the mean absolute classification error.

projection without considering parallel events) high improvement of results. For classification effectiveness it is +2.3% (totally 30.3%) and for classification effectiveness with tolerance it is +4% (65.2%).

However, the best results were reached for different value of parameter β in contrast to previous experiment (described in section 3.4). In this case it was $\beta = 8$ for classification and $\beta \in \{14, 15\}$ for classification with tolerance. For mean absolute classification error it was also $\beta = 15$. Projection of parallel events based on principle of multiplex or layered network achieved finally better results than projection used in common.

4.3.2 Influence of ageing factor

In our second experiment we used exponential metric for projection of events again with values 10 and 15 for parameter β. Graphically (see Figure 6) we evaluate effectiveness of classification for ageing factor given by parameter $t_{1/2}$ from 20 to 1500 days.

In our results we observe different progress of curves (compare curves a) and b) on Figure 6). Two "peaks" for curve b) are visible, the first one near time $t_{1/2} = 320$ and the second near time $t_{1/2} = 430$. Network projection with ageing of the ties provided better results of classification effectiveness. We reach results 32.4% (+2,1%) for direct classification and 67.0% (+1.8%) for classification with tolerance.

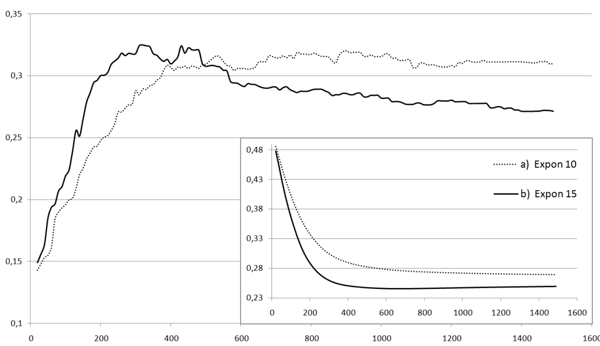

Figure 6: Dependency of classification effectiveness to the value $t_{1/2}$ for exponential weighting with a) $\beta = 10$; b) $\beta = 15$. In window: Dependency of mean absolute classification error to the value $t_{1/2}$.

At the end of this experiment we can point out that relations ageing allows evolution tracking of the collaboration ties (Figure 4) and it provided useful support for precise modeling of collaboration networks.

4.3.3 Influence of "real-time" customization of ageing factor

Results of this last experiment (see Figure. 7) shows small improvement of proposed method where ageing factor was customized by previous participations. Totally, in contrast to metric with constant ageing factor we achieved better results with classification effectiveness 32.8% (+0,4) and 67.2% (+0.2) for classification with tolerance. In both cases, the best results were reached for $t_{1/2} = 340$. Proposed metrics for customizing ageing factor by actual weight does not reach better results, so our assumption was not confirmed.

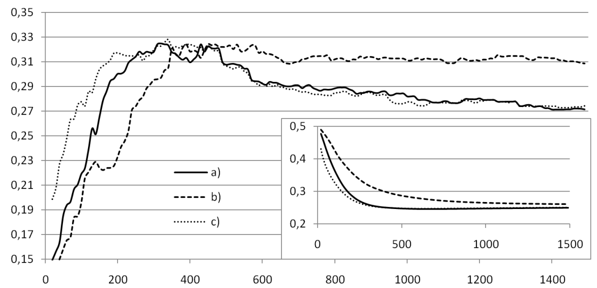

Figure 7: Dependency of classification effectiveness to the value $t_{1/2}$ for exponential weighting with $\beta = 15$ with a) constant ageing factor; b) ageing factor customized on relation weight; c) ageing factor customized on previous participations.

5. Case study - exploitation of information in collaboration social networks

Decomposition of the actors is process which performs partitioning actors into smaller groups or teams based on given rules. Some of the examples are: creation of school collectives at universities, support for teambuilding actions, challenges or partitioning participants of various free-time actions. Decomposition of the actors should achieve desired target with help of supporting data from the social networks. They store information about actor previous participations (or collaborations). It helps to decompose actors into these collectives in which actors have e.g. stronger relations inside then outside created collective. In "stronger" collectives actors have better working conditions and its work should by more effective. Opposite example of this is creation of such collectives where actors are unknown to each other in the same group/team. This is true for building new relations between them e.g. teambuilding actions or free-time activities.

5.1 Scenario for exploitation of information from social network

One of the possible examples of exploitation of social information we proposed for DAKCSN network which describes participation between members of free-time center. Members are partitioned into smaller groups or teams where they solve several tasks or they collaborate together. Before any actor attends its first action, he has

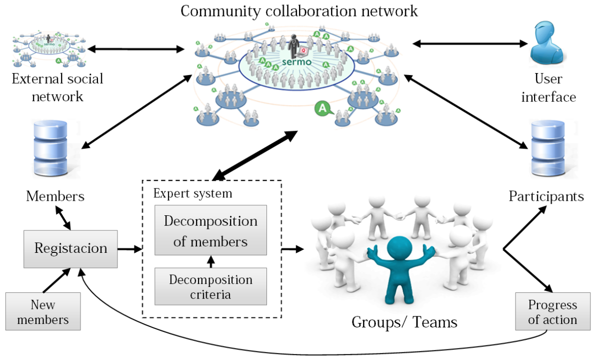

Figure 8: Scenario for exploitation of knowledge in Collaboration Social Networks.

to register as a new member and he also registers to any free-time action. After registration of all members who want to participate on specific action, the decomposition of actors is performed. Decomposition proceeds by several defined rules (see Appendix A.1). After execution of an action the database of participants (two-mode network) is updated. Data are next projected into one-mode network by proposed methods and all relations are updated[2]. One-mode network represents model of collaboration network which is used for further decomposition of actors. Scenario for exploitation of information from an social network is depicted on Figure 8:

5.2 Influence of aimed decomposition to network evolution

Application of aimed decomposition in DAKCSN network changes network global properties. We examine it by simulation of DAKCSN network evolution. Network contains information about past actions and decomposition of actors performed manually using only compositional attributes without considering structural data (see A.2). So we perform simulation trough all previous actions and we replace manual decomposition to automatic with considering compositional and all previous structural data. Collectives created automatically are characterized by weaker relations between collective members because we perform rules aimed for building new relations between network members. Simulated network evolution so in comparison to real network achieved better global characteristics like higher average node degree (actors know more others) in the network and higher count of relations, see Figure 9.

6. Conclusions

In section 2.1 we described base methods for projection of affiliation (two-mode) networks onto one-mode networks. In section 2.2 we identified three sub-processes of network projection. In sections 3 and 4 we described particular sub-processes of network projection in detail and we propose new approaches and improvements of existing methods. We performed three experiments, each one for different sub-process of network projection and we implemented improvements of proposed methods on DAKCSN and evaluate them on DAKCSN-Real dataset.

For first sub-process (section 3) we proposed exponential

[2]If new relations are established, they are added into network and if weight of some relations decreases by ageing below threshold ε, they are removed from the network.

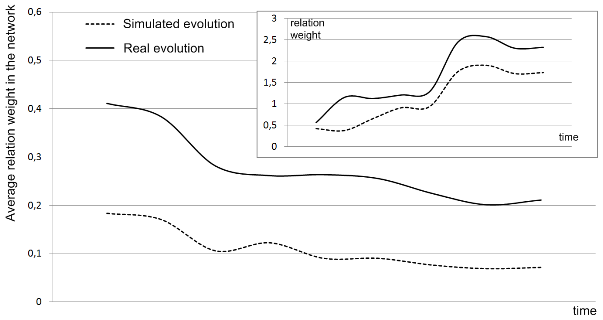

Figure 9: Influence of aimed decomposition to evolution: main - of average degree in the network; in window - of number of relation in the network.

(section 3.1) and sigmoid (3.2) weighting methods instead of binary or Newman weighting of collaborations. Both of proposed methods achieved better results than existing methods and exponential metric achieved little better results than sigmoid. We dealt with second and third sub-process of network projection in sections 4.1 and 4.2. For second sub-process we used principle of multiplex and layered networks and for third we proposed relation ageing instead of node ageing principle. This principle also improved our results and moreover it allows network analysis from new scopes like tracking collaborations over the time or creation and analyzing network snapshots in given time.

In our future work we will extend our experiments to different collaboration network datasets and we will compare usability of proposed methods on different types of collaboration networks.

Acknowledgements. The work presented in this paper was supported by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic under grant No. 1/0042/10 (50%). This work is also the result of the project implementation Development of Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Programme funded by the ERDF (50%).

Appendix

A. Data of community social network

DAKCSN community social network is a network of members (usually teenagers) of a non-profit organization dealing with organizing educative-pedagogic workshops for young people [4]. Network is created from historical data about participation of actors on free-time activities - *workshops*. Usually there are organized around 10 workshops annually with 150 - 700 participants on single workshop. The number of participants depends on workshop's type and duration. All participants of a single workshop are partitioned into smaller groups, usually with 8 - 12 members. Each group member cooperates with other group members and so there are established new social relations between group members.

A.1 Actor types in the network

Generally in DAKCSN network we recognize two main types (actor roles) of actors participating on actions - *participants* and *organizers*.

Participants - spending most of the time inside the group (i.e. they usually do not establish social relations outside the group). It is higher numerous type of actors (usually approx. 2/3 of all actors on action). Each of base participants is assigned into group of base participants without possibility to change it. They participate in this group during whole action and they establishe new relations with other group participants. Group assign process is performed by decomposition of all actors according to following rules:

- *Social relations* - group participants without or with minimal relations are selected into common group.
- *Balanced group size* - group size is within desired interval, usually for DAKCSN it is 10-12 members.
- *Actor's age* - actors of the same or the most similar age are assigned into same group.
- *Gender rate* - both genders, men and women are included in the group with the rate, which is most possible close to rate of man and women in the whole action.

Organizers - it is second main role of action participants who organize the whole action. They are partitioned into three sub roles:

- *Base organizers* - they work in teams by its choice and they are also assigned into a group of participants where they collaborate too.
- *Leaders of organizers team* - they manage base organizers and so they cooperate with them and they also cooperate with leaders of other organizers' teams.
- *Leader of participants group* - they role is selected by actor and he manages participants' group. Leaders of participants group cooperate together less than leaders of organizers.

A.2 Data in DAKCSN

Compositional attributes in DAKCSN data set are available for both, actors and events. Actors are described by attributes such as date of birth (age) and gender; as well as by geographical attributes - city or area of living. Events are described by their type. We can recognize two main types of events - events for base participants and events for organizers. Events for organizers are next categorized by their types of activity like registration, security or accommodation event.

Moreover, temporal attributes are available together with compositional attributes, e.g. start and end time of events and workshops. From these data we can derive several other attributes, such as "length of event" or "time of first visit" for particular actor. In our case it means the moment when the actor visited any event for the first time. This is one of the advantages of DAKCSN data set because we are able to track events in the time and recognize which events were organized in parallel and which sequentially. Also we are able to track participation of single actors on particular events.

This collaboration network can be expressed for single workshop by a bipartite graph as representation of two-mode network. The first mode is a set of actors, and the second mode is a set of events which affiliate the actors. We represent each group on the single workshop as a single event, so for one workshop we obtain several events. Additionally we added two more events for representation of cooperation between leaders of participants and leaders of organizers.

A.3 Testing (real) data

For testing purpose we obtained real data from inquiries - DAKCSN-Real. We randomly selected actors of the network with request to describe their collaboration relations. Each of the respondents expressed his/her collaboration strength to minimum thirty other network members. The structure of inquired network members was the following: members with potentially strong or strongest collaboration, members with potentially middle or weak collaboration and finally members without reciprocal collaboration. In such a way we have obtained real picture in form of a balanced collaboration network with 2278 ties between 828 distinct actors with their collaboration strength. Strength was partitioned into five nominal categories based on Likert five degree scale [9]: *No collaboration*, *Weak collaboration*, *Middle collaboration*, *Strong collaboration*, *Very strong collaboration.*

This real collaboration network created from data gathered in the inquiry was used as a reference network. As next we compared this reference network with equivalent parts of modeled networks gained from the DAKCSN system. We normalized weights of all networks into <0, 1> interval for correct evaluation and we also transformed five categories of collaboration strength into five regular intervals: <0, 0.2>, (0.2, 0.4> ... (0.8, 1>. By classification of weights of ties in modeled networks into these intervals we obtained precision of estimated weights by modeled networks.

References

[1] J. A. Bondy and U. S. R. Murty. *Graph Theory*. Springer, 2008.

[2] S. P. Borgatti. 2-mode concepts in social network analysis. *Encyclopedia of Complexity and System Science].*

[3] B. Cloteaux. Modeling affiliations in networks. *2010 Winter Simulation Conference*, pages 2958–2967, 2010.

[4] DAK. DAK - Collaboration Social Network. [online] at: `http://www.domcek.org/`, 2011. Data set of non-profit organization.

[5] A. Davis, B. B. Gardner, and M. R. Gardner. *Deep South. A social Anthropological Study of Caste and Class.* University of Chicago Press, Chicago, 1941.

[6] K. B. Hajra and P. Sen. Aging in citation networks. *Elsevier Science*, September 2004.

[7] J. T. Hamill, D. R. F. Decro, D. J. W. Chrissis, and D. R. F. Mills. Analysis of layered social networks. *IO Sphere*, pages 27–33, 2008.

[8] M. Latapy, C. Magnien, and N. D. Vecchio. Basic notions for the analysis of large two-mode networks. *Social Networks*, 30:31–48, 2008.

[9] R. Likert. A technique for the measurement of attitudes. *Archives of Psychology*, 140:1–55, 1932.

[10] M. Newman. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *The American Physical Society*, 64:64–70, 2001.

[11] M. Newman. Who is the best connected scientist? A study of scientific coauthorship networks. *Complex Networks*, 2004.

[12] T. Opsahl. Projecting two-mode networks onto weighted one-mode networks. [online] at: `http://toreopsahl.com/2009/05/01/projecting-two-mode-networks-onto-weighted-one-mode-networks/`, 2011.

[13] C. Stadtfeld and A. Geyer-Schulz. Analysing event stream dynamics in two mode networks. *INSNA - The Best Student Paper Award 2010*, 2010.

[14] G. Tutoky and J. Paralič. Modelovanie a analýza malej komunitnej sociálnej siete. In *5th Workshop on Intelligent and Knowledge oriented Techniloties*, Bratislava, 2010.

[15] G. Tutoky, M. Repka, and J. Paralič. Structural analysis of social groups in collaboration network. In *Faculty of Electrical Engineering and Informatics of the Technical University of Košice*, Košice, 2011.

[16] S. Wasserman and K. Faust. *Social Network Analysis*. Cambridge University Press, Cambridge, 1994.

[17] D. J. Watts and S. H. Strogats. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[18] H. Zhu, X. Wang, and J.-Y. Zhu. The effect of aging on network structure. *The American Physical Society*, 2003.

Selected Papers by the Author

G. Tutoky, J. Paralič. Time Base Modeling of Collaboration Social Networks. In proc. of the *3rd International Conference on Computational Collective Intelligence - Technologies and Applications*, Gdynia, Poland, Proc. LNCS, Springer. (in press), 2011.

G. Tutoky, J. Paralič. Modeling of Collaboration Social Network Including Temporal Attributes. In proc. of the *15th IEEE International Conference on Intelligent Engineering Systems*, Poprad, Slovakia, 2011.

G. Tutoky, J. Paralič. Collaboration Social Network with Temporal Attributes. Article selected into the book from *15th IEEE International Conference on Intelligent Engineering Systems*, Proc. LNCS, Springer. (will press on end of 2011), 2011.

G. Tutoky, J. Paralič. Projection of Affiliation Network and Modifications of Positional Analysis. In proc. of the *5th Workshop on Intelligent and Knowledge Oriented Technologies 2010*, Bratislava, Slovakia, 2010.

M. Repka, G. Tutoky. Projection of Affiliation Network and Modifications of Positional Analysis. In proc. of the *Conference of Young Researchers 2011*, Herľany, Slovakia, 2010.

G. Tutoky, M. Repka. Structural analysis of social groups in collaboration network. In proc of the *Electrical Engeneering and Informatics II*, Technical University of Košice, Košice, Slovakia, 2010.

G. Tutoky, J. Paralič. Modelovanie a analýza malej komunitnej sociálnej siete. In proc of the *5th Workshop on Intelligent and Knowledge Oriented Technologies*, Bratislava, Slovakia, 2010.

J. Paralič & K. Furdík & G. Tutoky. & P. Bednár & M. Sarnovský & P. Butka & F. Babič. Dolovanie znalostí z textov. Equilibria, 183s. ISBN: 978-80-89284-62-7, 2010.

G. Tutoky, J. Paralič. Rozpoznávanie pomenovaných entít v medicínskych textoch. In proc of the *Znalosti 2010*, Jindřichuv Hradec, Czech Republic, 2010.