Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Group Formation in eLearning-enabled Online Social Networks

Steffen Brauer

Seminar WiSe 2012/13

Fakultät Technik und Informatik
Department Informatik

Faculty of Engineering and Computer Science
Department of Computer Science

Steffen Brauer

Seminar WiSe 2012/13

Eingereicht am: 28. Februar 2013

Contents

1 Introduction

Online Social Networks (OSN) stimulate their users to socialize with friends and communicate to each other. Discussions in groups are user-triggered and do not need a moderator or facilitator. OSNs enjoy an overwhelming popularity among students.

eLearning Content Management Systems (LCMS) allow physically distributed users to access structured content and to collaborate via inter-group communication on learning topics. Modern LCM-systems organize content in eLearning objects that interrelate to form an instructional or semantic network [1]. Usually they are bound to an instructor who creates groups, analyzes course results, and tracks learning progress. The use of LCMSs is commonly limited to dedicated courses or schools.

Our work tries to open the learning process and the building of learning groups to become part of social Internet eco system. We concentrate on integrating an OSN with an LCMS, thereby removing its dependency on an instructor. Such an eLearning-enabled OSN allows users to self-pace learning in topics of personal interest and teams of personal choice. The removal of an instructor in eLearning scenarios leads to the following challenges in designing the OSN [2]:

1. How to stimulate a team building process that is effective for learners?
2. How to provide access to the relevant content for a learning group?
3. How to facilitate a consistent learning progress, include feedback and corrective actions?

This paper addresses the first question. Learning in groups creates motivation for a user through the ability to compare to each other, provided the group is well formed. There are many possible factors that can influence the quality of a learning group. Often criteria like knowledge and learning style are taken into account, but via an OSN it is also possible to account for social relationships between the users when building groups. In this context, possible metrics are the weights of the edges between two users for indicating former collaboration [3], or creating a representation of trust between users [4]. In our approach, we concentrate on user's availability, knowledge, learning style and the group density in the social network when forming a group for collaborative learning.

To access the problem of group formation more easily , our approach is divided in two parts. First we browse the social network and try to find a minimal number of suitable candidates for the formation of a group, which an initiator shaped on a chosen topic. Based on the candidates, the second part tries to optimize candidate constellation for a successful group learning experience. Both steps are grounded on metrics that are calculated from user configuration and statistics in the underlying Online Social Network.

This paper provides an outline for the group formation approach by addressing the objectives in Section 2 and risks in Section 4. In Section 3 the Group Formation Engine is introduced. An in-deep presentation of current work can be found in [5]. Finally, the paper ends with a conclusion and an outlook on our future research.

2 Problem Statement and Objectives

The challenge of forming a group for an effective mutual learning process lies in finding those people that are not only interested in the same subject area, but are thematically at eye level and match in relevant social dimensions. It is our aim to harvest appropriate candidates from an augmented OSN.

Social network sites can be considered as an undirected graph with vertices built from users and edges that represent the relationship between them [6]. In this perspective, the problem of team building in OSNs is forumlated by Roreger and Schmidt [2] as finding a subgraph of the full social network that fulfills the following conditions:

1. Each user is motivated to collaboratively learn on a certain topic
2. The learning style of a user is appropriate to form a balanced group
3. The background on the topic is compatible among group members

Condition 1 is an intrinsic motivation of a person. The recognition of the motivation can be implemented in different manners, but is not demanding. A simple approach is to set a flag on one's personal profile. This flag initially indicates a person's interest in collaboration. Later on the usage of the eLearning system is monitored to track a user's motivation.

Condition 2 requires information on a user's learning style as well as a mechanism to find a subgraph of the OSN-graph which is balanced in terms of learning. This is achieved by grouping people who learn in a similar learning style.

Condition 3 is about finding groups of learners with a common knowledge base. Since information in OSN is widely spread in different, non-standardized formats, automatic knowledge estimation and representation through analyzing semantics is an advanced topic.

2.1 Learning Style Assessment

A natural aspect to evaluate collaborative potentials in learning is harmony in learning style. Learning style models, though, are sometimes criticised according to their reliability, validity, and implication for pedagogy. Coffield et al. [7] presented a review of learning styles and conclude that there is a lack of theoretical coherence and a common framework. Nevertheless, the use of learning style in eLearning application for selecting a certain content representation can improve the learning experience [8, 9, 10].

We employ learning style in agreement with Felder and Silverman's theory (FST) [11]. This work is widely accepted as a standard way to assess learning styles. A key feature of this theory is that it does not try to force a user into one specific category of leaning, but variably assigns preferences to a learner in the four predefined dimensions:

- "Active or Reflective" (Processing)
- "Visual or Verbal" (Input)
- "Sensing or Intuitive" (Perception)
- "Sequential or Global" (Understanding)

In each of these dimensions, the user can have three different strengths, i.e., fairly well balanced, moderate preference and a very strong preference. Measuring learning styles within an eLearning application has many advantages. We want to use this as an instrument for customizing search, select a presentation of content and use it in the group formation to create groups with a common learning style.

2.2 Knowledge Representation

In the Semantic Web, knowledge is commonly represented in ontologies. However, the creation and maintenance of ontologies requires experts on the subject and lacks flexibility. Reasoning based on ontologies is in addition complex and often slow.

Our design of a platform represents knowledge by tags. In contrast to ontologies, we want to use a lightweight approach that combines knowledge annotations for users, content and topics. This largely increases flexibility: While ontologies are only able to represent a Web of a special topic, tags can jointly describe content, the competence of a user, or the context and style of a content object [4]. After finishing a topic the tags assigned to the topic are transferred to the user. Each tag, which is assigned to a user, is weighted by an activity index. This index represents the relevance of a user per tag and is accumulated during the history in the system as a normalized exponential average. To match a topic with possible group members, each topic has a tag vector with weights that encode the relative relevance.

3 Approach

Based on the objectives defined in the previous section, now we can introduce our group formation engine. This Section starts with an overview of the eLearning-enabled OSN. Based on the characteristics of the network, we describe our approach to build learning groups and give a short summary of the evaluation.

3.1 eLearning-enabled OSN

While the eLearning content management system hylOs is used as a content repository, our extensions target at the Open Source Online Social Network Diaspora to create our eLeanring-enabled OSN. The communication features of Diaspora will be extended by learning structure objects. The entire network is modelled by an undirected graph with different types of vertices. These types capture all content kinds, user profiles, groups and topics. Relations (links) are typed accordingly. This unified approach, cf. [12], adds many implicit relations to the network. In this way it is possible to find users, who have object relations in common, but no personal interconnect. It also enables algorithms to measure the strength in connectivity of two vertices by accounting for shared neighbours or distinct paths that connect them.

The user-centric nature of the OSN positions user objects in the key role among vertex types. We extend the given profile of the original OSN by including an availability flag of a user, and by encoding the knowledge and learning style. Another vertex type is the topic object. It describes a task or a field of work. Edited and managed by a user, it also includes a definition of the desired knowledge and a number of required collaborators. To simplify the search of relevant content for a topics, the topic vertex can be connected with a content object. These objects can represent any kind of content that is managed by the same user. To associate users with a topic they are working on, a group object is needed. If a group is created, all members subscribed to the chosen topic connect to the group objects.

3.2 Group Formation Engine

The objective of our group formation engine is an automated evaluation of proper learning groups. For this, we now discuss algorithms that retrieve suitable candidates for performing a specific learning task and suggest collaborative groups to any user who initiates it.

Our algorithms define metrics that target at the requirements discussed in the previous sections. Based on learning style recognition and knowledge estimation data, these metrics calculate proximity in the context of learning between members of an eLearning-enabled OSN. The group formation engine proposes a set of users to collaborate with each other based on these distances between learners.

To start group formation, a user decides to initiate collaborative work on a topic. Initially, topics are user-defined, but can be selected later. The system starts searching for candidates and suggest different group constellations. Now the initiator selects a preferred group and invites all group members. If all members agree to work in this group, they can immediately begin collaboration on the topic. After the task is finished, the group closes.

3.2.1 Metrics in the Social Learning Space

According to the objectives defined in Section **??**, we need to quantify the user's availability, learning style and competence level, but also want to account for social proximity. The availability of a user u is simply modelled by the function $A(u)$ which returns $true$, if user a is available, or $false$ otherwise. We now define the other quantities as distance metrics in the social learning space. Notably, all metrics are normalized to range from zero to one.

The learning style is represented as a vector $L(u)$ with an entry for each dimension of the Felder and Silverman Theory. Possible values are 1, 0 and -1 indicating a positive or neutral or negative characteristic in each category. The learning style distance $D_L(u, v)$ between two users is evaluated as the Euclidean distance between the vectors and normalized by its maximum possible value 8:

$$D_L(u, v) = \frac{1}{8} \sum_{i=0}^{4} |L_i(u) - L_i(v)| \tag{1}$$

As introduced in Section 2.2, our knowledge model is build on tags, which are assigned to a chosen topic for each user. A node in the network, a user, a group, or a subject, holds a list of tags and a vector which represents the weights of each tag in its context. For example, a user that initiates a topic t, i.e., a group subject to team building, selects a list of tags $t = \{\tau_i\}_i$ and assigns individual weights $W(\tau_i)$ to them.

Each user likewise carries a list of tags, which are acquired from interactive work on topics jointly with an activity index (see Section 2.2). The corresponding activity index $I_A(u, \tau)$ is a normalized weight obtained from exponential averaging.

To calculate the distance D_K in knowledge between a topic and a user, the first step is to match the user's tags to the topic. (A tag of a topic not present at a user gets an activity index of zero assigned). After tags match, we can calculate the correlation of the topic and the user tags as the scalar product of the weight vector of the topic $W(\tau)$ and the activity vector of a user $I_A(u, \tau)$:

$$D_K(t, u) = 1 - \langle W, I_A \rangle \equiv 1 - \sum_{i=0}^{n} W(\tau_i) \cdot I_A(u, \tau_i) \tag{2}$$

This value indicates, how the displayed knowledge of a user correlates to a certain topic. Note that the normalized scalar product is one, if the user admits full activity in the topic, and zero, if user's activities do not overlap with the topic.

The total distance D between an initiating node u and a chosen topic t with a possible candidate v is calculated as the weighted sum of these two parts

$$D(u, v, t) = D_K(v, t) + D_L(u, v). \tag{3}$$

This 'learning distance' $D(u, v, t)$ shall be small enough and will serve as our selection function for candidates.

When considering teams, we want to take additional advantage of the social dimension. For measuring the distance of two users in a social graph, we first preselect a maximal diameter δ_{max} of our graph under consideration. Due to the small world property, this diameter is usually very small (e.g., three to five). The social distance D_S between two users is then defined as the normalized shortest hop-length,

$$D_S(u, v) = min_{paths} \left\{ \frac{\delta_{path}(u, v)}{\delta_{max}} \right\}, \tag{4}$$

where $\delta_{path}(u, v)$ denotes the hop-length of a path from u to v, and the minimum is taken over all such paths.

3.2.2 Candidate Selection

The first step of our group formation approach is the candidate selection. Its task is to extract possible group members from the underlying social network. To reduce the complexity of group formation, it is necessary to select a small set of well suiting user-nodes. Starting at the initiator, the network is searched for nodes with a common learning style and knowledge base as evaluated by Equation 3.

The choice of the search algorithm is essential for the group formation process. Because several algorithms optimized to social networks try to find special nodes, the group distance in the social network is here relevant.

To reduce the complexity of the candidates selection, it can be parametrized with the maximal number of candidates and a threshold, which determines whether a node is added to the candidate set. These parameters determine the quality and complexity of the result. If the threshold is high, the candidates are near to the initiator, but may have a higher distance in the sense of learning style and knowledge. For a low threshold, the search algorithm will select nodes that have a higher distance in the social network, but are closer in the sense of learning style and knowledge. By choosing a low threshold the performance decreases.

The position of a user-vertex is not used in the candidate selection, because in this phase of the algorithm the candidates are a loosely coupled set and no statements can be made about group membership. So it remains open what the final group density will be. It is noteworthy that the initiator plays no special role and could be the least connected part of the group.

To select the best-suited search algorithm it is necessary to take the overall requirements into account. By considering the density of group members in the social network, a team with experts on their topic at an equal learning style, but with a low density in the network does not satisfy the requirements. Also this team configuration would have high computational cost. Three different search algorithms are selected based on the evaluation of Zhang and Ackerman [13] and assumptions concerning the distance of the searched nodes in the social network.

Breath First Search (BFS) is a classic way to traverse a graph. Starting from the initiator, BFS will probably find the nearest candidates, because it traverses the social network with stepwise increasing distance.

Random Walk (RW) introduced by Adamic and Adar[14], traverses the social network by random paths. In contrast to BFS, the distance to the initiator in Random

Walk increases very fast. This could lead to a selection of candidates who have a high distance to the initiating node. This phenomenon can be reduced by restarts.

Best Connected Search (BCS) performs well at networks with a power-law distribution of nodal degrees. The strategy is to select nodes by the number of neighbors [15].

3.2.3 Group Optimization

After a set of candidates has been selected, the next step is to find a group constellation that is dense enough to be recommended to the initiator. To achieve this goal, we generate the set M of all candidate groups that satisfy given constraints on group sizes. We then optimize our metrics defined in Section **??** according to the entire group.

In detail, we consider each suitable group $g \in M$ for a given topic t and want to minimize the overall group distance

$$D_G(t,g) = \binom{|g|}{2}^{-1} \sum_{u,v \in g} \{D(u,v,t) + D_S(u,v)\}\,. \tag{5}$$

Here we include the social distance to arrive at a socially balanced group constellation. Note that $D_G(t,g)$ is renormalized and attains values between zero and three.

This minimization problem is equivalent to maximizing the following group fitness function over all $g \in M$ that meet the constraints on group size:

$$G_{Fit}(t,g) = 3 - D_G(t,g) \tag{6}$$

The fitness function evaluates the overall density of a group and accounts for the dimensions of learning, background knowledge and social proximity.

For practical evaluations, we need to focus on computational cost. Ranking all possible group constellation with G_{Fit} works fine for a small number of candidates. If the number increases, too many constellation arise for a full computation. We treat this scalability problem by using Genetic algorithms.

Mapped to a Genetic algorithm, a set of team configurations is represented as a population of chromosomes. Each chromosome is a group with users represented as genes. In each generation, crossover and mutation operations are performed on the population. A crossover population splits two chromosomes and exchanges the parts. A mutation exchanges only one gene in the chromosome with another. Applying this to our approach another user is selected from the candidate set. When the operations finish, the fitness of all chromosomes is evaluated and the best are selected for the next generation.

After sufficiently many generations have been run, the best group constellation is recommended to the user, who can now send invitations for joining the group to the selected candidates.

3.3 Evaluation

As a pre-study to real-world deployment of our OSN, we have performed an evaluation of our group formation approach. This task raised the problem of proper test data. There are several models for the generation of social networks with real world features. We built the base structure of the social network by using the Forest Fire Model [16] as proposed by Leskovec et al. This model reflects the characteristic features of a network and is popular in the literature, because it creates networks that closely resemble real measurements [17]. eLearning content management systemFor our evaluation, we generate a graph with 1000 vertices and 31522 edges. In this test network, only user vertices are created, which eases the evaluation of the group formation process. We omit simplifying effects of the unified approach of our eLearning-enabled OSN.

The next step for generating a test network is the assignment of user data. The easiest way of achieving this would be a random distribution of values to each vertex. But this could lead to a test network with unrealistic characteristics, as was indicated by findings of Derntl and Graf [18]. The authors started from a blog as a learning diary to a course and tried to identify correlations between the blogging behaviour and the learning style of the students. By comparing the blogging behaviour and the active reflective dimension, they found a correlation to the number of blog posts. Active learners tend to write more blog posts than reflective. On the other hand, reflective learners read more posts than active. In addition active learners tend to follow the chart of rated blog posts because of their social orientation. These findings indicate a correlation between the degree of a user in the social network and the value in the active resp. reflective dimension of the learning style. How learning style correlates with social network characteristics will be a part of our future research.

Another problem in assigning learning styles is to choose the distribution of dimension values. Felder and Spurlin [19] summarize the result of different studies and an average distribution can be used to assign the dimension values.

Besides the distribution of learning style, we want to assign the tags of a user based on empirical data. There are several measurements on tags. Sigurbjörnsson and van Zwol

2008 [20] evaluated the tag distribution on flickr and found a power-law distribution on unique tags and photos per user. Rodrigues et al [21] evaluate on question answering sites.

To achieve a realistic input to our Knowledge Rank calculation, we evaluate data available from Stack Overflow. Stack Overflow is a Question and Answer site for programmers with 1.2 million users and 3.3 million questions asked. Within the Stack Overflow platform, posts like questions, answers and comments can be tagged by users. To employ these tags as a knowledge representation for our network, we measure the distributions of unique tags and their assignments to users through their answer posts.

Our candidate selection is evaluated by focusing on two aspects. First to identify which search algorithm selects the best candidates compared by their fitness, and second to learn how much nodes are visited while selecting the given number of candidates.

In the final evaluation of the group formation progress, we try to answer two questions: (i) How do optimal group selection functions differ from heuristics? (ii) How is the correlation between the number of potential candidates and actual group sizes. We analyse these questions with the help of the fitness function. In our setting, a high fitness value indicates that the group formation process led to a near-optimal group.

4 Risks

Based on our appraoch, we can identify risks which have to be taken into account by preforming the evaluation and statements on its results.

4.1 Quantification of Users

One important risk is the quantification of user characteristics. This is a general problem by systems, which try to analyse and use characteristics of real persons. We attempt to minimize this risks by using well studied Felder and Silverman model for learning style and a flexible model for knowledge representation.

4.2 Evaluation on Synthetic Data

Focusing on the evaluation of our approach, the quality of test data is relevant. To create a realistic test bed, we use empirical data presented in the literature and preform own measurements on Stack Overflow.

5 Conclusion

This paper introduces a central building block on the path to eLearning-enhanced social networking. We discussed the central research questions in this field, and focused on the problem of group formation within a knowledge-aware social network.

We presented a multi-steo group formation approach for an eLearning-enabled Online Social Network. Based on a unified social network, which includes all functional objects like users, content, groups and topics, a roadmap to searching suitable candidates was derived. A user is selected as a candidate with respect to his learning style, knowledge, and social proximity. To represent the learning style of a user, we employed the theory of Felder & Silvermann. The knowledge and competence of a user was represented by tags with activity weights assigned. To compare the requested knowledge for a given topic with the knowledge provided by a user, we calculated a knowledge rank build on a importance vector of the topic and the activity index vector of a user.

The efficiency to search for suitable candidates in a real-world social network is of vital practical importance. We implement three search strategies, Breath First Search, Random Walk Search and Best Connected Search. Based on the resulting candidate sets, we applied genetic algorithms to find the best group constellations using a group fitness function that includes (i) the distance between the group members in learning style, (ii) the level of knowledge ranks, and (iii) the group density in the social network.

In our future work, we will improve the social metric in group optimization and expand the evaluation to a multi entity network to evaluate the impact of the unified social network.

References

[1] M. Engelhardt and T. C. Schmidt, “Semantic Linking – a Context-Based Approach to Interactivity in Hypermedia,” in *Berliner XML Tage 2003*, R. Tolksdorf and R. Eckstein, Eds., Humboldt Universität zu Berlin, September 2003, pp. 55–66.

[2] H. Roreger and T. C. Schmidt, “Socialize Online Learning: Why we should Integrate Learning Content Management with Online Social Networks,” in *Proc. of IEEE Intern. Conf. on Pervasive Computing and Communication (PerCom), Workshop PerEL*. Piscataway, NJ, USA: IEEE Press, March 2012, pp. 685–690.

[3] C. Dorn, F. Skopik, D. Schall, and S. Dustdar, “Interaction Mining and Skill-dependent Recommendations for Multi-objective Team Composition,” *Data & Knowledge Engineering*, vol. 70, pp. 866–891, 2011.

[4] J. Vassileva, “Toward Social Learning Environments,” *Learning Technologies, IEEE Transactions on*, vol. 1, no. 4, pp. 199 –214, oct.-dec. 2008.

[5] S. Brauer and T. C. Schmidt, “Group Formation in eLearning-enabled Online Social Networks,” in *Proc. of the International Conference Interactive Computer aided Learning (ICL’12)*, M. E. Auer, Ed. Piscataway, NJ, USA: IEEEPress, Sep. 2012.

[6] I. Foudalis, K. Jain, C. Papadimitriou, and M. Sideri, “Modeling Social Networks through User Background and Behavior,” in *Proc. of the 8th int. conf. on Algorithms and models for the web graph (WAW’11)*, 2011, pp. 85–102.

[7] F. Coffield, D. Moseley, E. Hall, and K. Ecclestone, “Should we be using learning styles? What research has to say to practice,” The Learning and Skills Research Center, Tech. Rep., 2004.

[8] M. Budhu, “Interactive web-based learning using interactive multimedia simulations,” 2002.

[9] C.-I. Peña, J.-L. Marzo, and J.-L. Rosa, “Intelligent Agents in a Teaching and Learning Environment on the Web,” in *Proc. of International Conference on Advanced Learning Technologies (ICALT2002)*, 2002, pp. 21–27.

[10] N. Stash and P. D. Bra, “Incorporating Cognitive Styles in AHA! (The Adaptive Hypermedia Architecture),” in *Proc. of the IASTED International Conference Web-Based Education*, 2 2004, pp. 378–383.

[11] R. Felder and L. Silverman, “Learning and teaching styles in engineering education,” *Engineering education*, vol. 78, no. 7, pp. 674–681, 1988.

[12] E. Amitay, D. Carmel, N. Har’El, S. Ofek-Koifman, A. Soffer, S. Yogev, and N. Golbandi, “Social search and discovery using a unified approach,” in *Proceedings of the 20th ACM conference on Hypertext and hypermedia*, ser. HT ’09. New York, NY, USA: ACM, 2009, pp. 199–208.

[13] J. Zhang and M. S. Ackerman, “Searching for expertise in social networks: a simulation of potential strategies,” in *Pro. of the 2005 int. ACM SIGGROUP conf. on Supporting group work*, ser. GROUP ’05, 2005, pp. 71–80.

[14] L. A. Adamic and E. Adar, “How To Search a Social Network,” *Social Networks*, vol. 27, 2005.

[15] S. Arafeh and D. Levin, “The Digital Disconnect: The Widening Gap Between Internet-Savvy Students and Their Schools,” in *Proceedings of Society for Information Technology and Teacher Education International Conference 2003*, C. C. et al., Ed. Chesapeake, VA: AACE, 2003, pp. 1002–1007.

[16] J. Leskovec, J. Kleinberg, and C. Faloutsos., “Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations,” in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2005, pp. 177–187.

[17] A. Sala, L. Cao, C. Wilson, R. Zablit, H. Zheng, and B. Y. Zhao, “Measurement-calibrated graph models for social network experiments,” in *Proceedings of the 19th international conference on World wide web*, ser. WWW ’10. New York, NY, USA: ACM, 2010, pp. 861–870.

[18] M. Derntl and S. Graf, “Impact of Learning Styles on Student Blogging Behavior,” in *Ninth IEEE International Conference on Advanced Learning Technologies (ICALT)*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 369–373.

[19] R. M. Felder and J. Spurlin, “Applications, Reliability and Validity of the Index of Learning Styles,” *Int. J. Engng Ed.*, vol. 21, no. 1, pp. 103–112, 2005.

[20] B. Sigurbjörnsson and R. van Zwol, “Flickr tag recommendation based on collective knowledge,” in *Proceedings of the 17th international conference on World Wide Web*, ser. WWW ’08. New York, NY, USA: ACM, 2008, pp. 327–336.

[21] E. M. Rodrigues, N. Milic-Frayling, and B. Fortuna, "Social Tagging Behaviour in Community-Driven Question Answering," in *Proc. of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 112–119.