User-Triggered Structural Changes in OSN-Alike Distributed Content Networks

Hauke Coltzau and Mario Kubek

Abstract We continue the evaluation of our model to describe regularly occurring phenomena in online social networks. The influence of link removals triggered by dislikes on the network structure is analyzed. Additionally, two scenarios for link replacement are investigated. It is shown that structured link replacement in the local neighborhood has acceptable impact on the graph structure and substantial benefits for the average like rate.

Keywords Online social networks · Modeling

1 Introduction

The remarkable sizes of today's online social networks (OSNs) allow for a unique practical view on the development of global behavior patterns in user-controlled decentralized systems. Interesting characteristics of these systems are the distributions of relationships between users, the diameter of the structure as well as the perseverative occurrence of superpopular contents.

In previous works, a simple model for the distribution of contents in OSNs was discussed. It could be shown that the occurrence of superpopular contents is very robust against parameter changes in this model [4]. Ongoing work will focus on the influence of network changes on the distribution and occurrence of superpopular contents. The work described in this article connects these two topics, the focus therefore lies on the enhancement of the basic model to integrate user triggered changes in the network structure based on the evaluation of recommendations received from other users.

H. Coltzau (✉) · M. Kubek
Fernuniversität in Hagen, Hagen, Germany
e-mail: hauke.coltzau@fernuni-hagen.de

M. Kubek
e-mail: mario.kubek@fernuni-hagen.de

P. Meesad et al. (eds.), *Recent Advances in Information and Communication Technology 2016*, Advances in Intelligent Systems and Computing 463,
DOI 10.1007/978-3-319-40415-8_2

The remaining article is organized as follows: In Sect. 2, an overviewing extract of existing works on OSN analysis and modeling is given. Our model is described in Sect. 3. Simulations in Sect. 4 discuss the results for three different link-adding scenarios.

2 Related Work

A remarkable amount of scientific works already exists in the field of online social network analysis, covering various fields like computer science and mathematics, but also physics and sociology. Almost all relevant online social networks already have been analyzed with regard to information flow, network structure and user behaviors.

Kleinberg [8], Watts and Strogatz [13], and Albert [1] and Barabási [2] independently have described or predicted reoccuring structural properties relying only on knowledge of networks with much smaller scale. They proposed different models for the evolution of large graphs, but merely focused on the distribution of the nodes degrees, their connectivity and the graph diameters. The distribution of contents was not part of their analysis, neither were the influences of dynamic user-triggered changes.

Cha et al. [3] found popularity growth patterns for pictures in the Flickr OSN. Lerman and Gosh [9] analyzed news spreading in Digg and Twitter from a user activity point-of-view. They deduce that in both platforms, news spread in an almost deterministic way similar for both platforms. Doerr et al. [5] confirm that information in social networks is propagated very quickly due to influential users. Zhang et al. [14] developed mechanisms to efficiently identify these influential users in a mobile online social network environment.

As for the analysis of social network structures, Schiöberg et al. [10] tracked the growth of Google+. Viswanath et al. [12] found that most of the links between users in Facebook are only seldomly used, while few links show a high level of activity. Additionally, it could be shown that the activity on a single link remarkably decreases over time.

Only few works exist regarding abstract modeling of online social networks. Kleinberg [8], Easly [6], as well as Watts and Strogatz [13] describe core principles on the efficiency of social distributed systems. But again, the perspective of their works is descriptive and does not discuss individual actions of users and their influence on the networks. Tovionen et al. [11] propose a network creation model that resembles some existing social networks e.g. in terms of degree correlations and clustering. Jamali et al. [7] derived a user rating-behavior model especially of the Flickr network but only with regard to the dynamics of the rating system and not as a general model.

3 Model

The model is an extension of the still unpublished works of Böhme and Airmarn. Let there be a set of users U and a set of files C that represent content distributed over the network. For now, we let C be $\in \mathbb{N}$, although future works will focus on more realistic document corpora. The dissimilarity $\delta(c_i, c_j)$ between any two contents $c_i, c_j \in C$ shall simply be their Euclidean distance. As usual, the structure of the distributed system is defined by the neighborhood relationships of their users. For each user, a neighborhood $N(u) \in U$ is defined, containing all users, a directed connection exists to.

Users can recommend content they already know to other users in their neighborhood. For this purpose, each user $u \in U$ maintains a recommendation stack $R(u)$. To recommend a content to a neighbor $n \in N(u)$, a user pushes the content to $R(n)$. In a subsequent step, the target user of a recommendation can either accept ("like"), ignore or actively reject ("dislike") the recommendation. To evaluate recommendations received from other users, each user defines a set of interest areas I_u containing tuples (c_i, d) with $i_c \in C, d \in \mathbb{R}^+$ during initialization, where d is a system-wide parameter and all c_i are chosen randomly from C. In the same way, a set of avoided topics $A_u = \{(c_a \in C, d \in R)\}$ is defined.

An element $c \in R(u)$ is evaluated using a decision function $f : C \rightarrow -1, 1$ that maps c to a value from -1.0 (dislike) over 0.0 (ignore) to 1.0 (like) according to the minimum dissimilarity to the user's interest and avoidance areas. Recommendations that are within a distance of d to any of the users interests $c_i \in I$ are mapped to 1.0 (like). All distances between d and $2 \cdot d$ are linearly mapped to a value in [1.0..0.0] and all distances above $2 \cdot d$ are mapped to 0.0. In the same way, the dissimilarity to the closest avoided topic is taken to map the recommendation to a dislike-value in [-1.0..0.0] If a content is within the radius of $2 \cdot d$ of both an interest area and an avoidance area, the results of the evaluation functions for both areas are added.

The elements, a user likes (i.e., those, who evaluate to 1.0), are put to the users *like-stack* $L(u)$, which serves as memory with a maximum size of $k_u \in \mathbb{N}^+$, where a new element will replace the oldest one in case the maximum size is reached. Each element in $L(u)$ can be recommended to the user's neighbors (Table 1).

Initially, each user's like stack $L(u)$ is randomly filled with elements from the user's interest areas. Each timestep, a user can only process the top element from the recommendation stack $R(u)$ and then decide, if the element is accepted and transferred to $L(u)$, dropped (ignored) or rejected. In the same timestep, the user may also recommend one element from $L(u)$ as described above.

Each user u keeps track of evaluation results received from and given to other users. For each user $e \in U$, a recommendation was given to or received from, a limited set $M(u)_e$ exists containing the last m recommendation results for/from user e to/from user u. Should the average value of these m entries in $M(u)_e$ fall below a system wide threshold Δ, the user u will initiate a removal of any existing connection $u \rightarrow e$ or $e \rightarrow u$ between itself and e. As long as $M(u)_e$ does not yet contain of m elements, no action is performed at all.

Table 1 Model parameters

$\|U\|$	Number of users in the network. Here, this number is fixed to 1000
$\|C\|$	Number of available distinguishable contents, also fixed to 1000 here
p_{int}	The probability for each $c \in C$ that c is of interest for $u \in U$
p_{avoid}	The probability for each $c \in C$ that c is to be avoided by $u \in U$
d	The maximum dissimilarity to any of a user's interest (avoidance) areas, recommended content may, to be liked (disliked) by the user
k	The size of the recommendation memory of a user, i.e. the maximum for $\|L(u)\|$
Δ	Threshold for average result for the evaluation of recommendations to remove a connection between two users
r	The depth of the neighborhood search for best-fitting neighbors

4 Simulation and Results

We investigate three different scenarios for how users deal with the removal of a link to another user:

1. In a *remove-only* scenario, no further action occurs, when a link is removed.
2. In the second scenario, a removed connection is replaced with a *random link* to any user in the network.
3. In the *best-fitting neighbors* scenario, the user parses through their neighborhood to find a neighbor, whose interest areas differ as little as possible from the interest area of the user itself. The deepness of the search is given by the simulation parameter r.

All simulations displayed in this article have been set up with the same number of users (1000) and contents (1000). In all cases, the graph properties of the user connection graph where tracked as well as the average value over all user's recommendation memories, which could be interpreted as average acceptance over all recommendations.

4.1 Scenario 1—Remove Only

The underlying initial network is a Watts/Strogatz with $k = 4$ and $\beta = 5\,\%$, hence the clustering coefficient is rather high and the average path length relatively low in the beginning of the simulation. The clustering coefficient decreases after the first users begin to remove links, the time of which being almost independent from the chosen threshold Δ (see Fig. 1a). At the same time, the average path length starts to increase (Fig. 1b). Nevertheless, after a rather short period of time, both graph properties stablilize at a value depending on Δ, where higher values for Δ lead to a

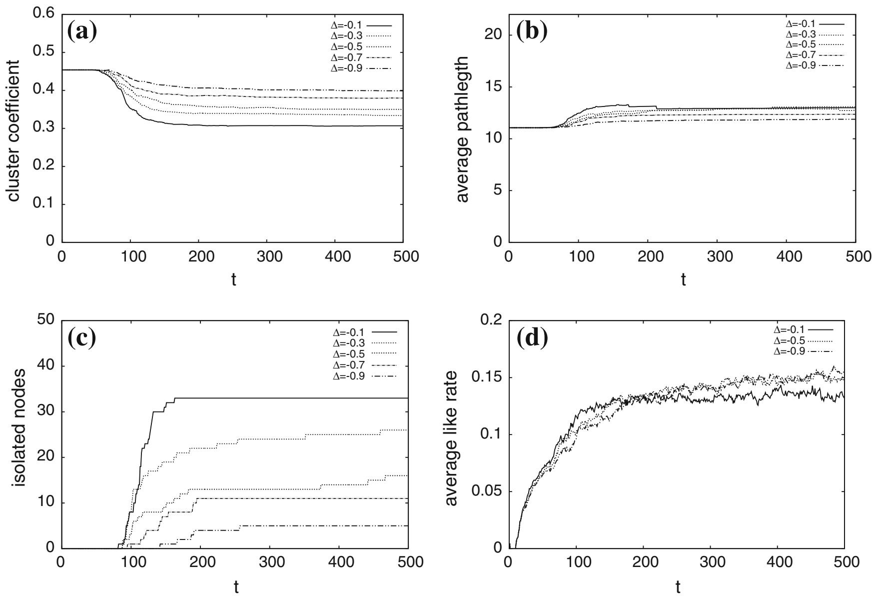


Fig. 1 Scenario 1—remove links only. **a** Clustering coefficient. **b** Average path length. **c** Isolated nodes. **d** Average like rate

higher clustering coefficient and lower average path length, because fewer links are actually removed.

By removing links, the user graph becomes disconnected. Some nodes do not have any incoming or outgoing neighbours with sufficiently overlapping interest areas any more. These nodes become isolated over time (see Fig. 1c). Not surprisingly, the number of isolated nodes also depends on the value of Δ, such that higher values for Δ result in lower numbers of isolated nodes.

As a general result for this scenario, it can be concluded that removing links depending on the average like value of recommendations does influence the connectivity of the graph as well as its clustering coefficient negatively. Nevertheless, both graph properties converge after a short period of time.

4.2 Scenario 2—Random Links

Adding random links after deletion of neighborhood relations transforms the user graph into a mixture between a Watts/Strogatz small world and an Erdös-Rényi random graph. Therefore, it is not surprising that the clustering coefficient decreases remarkably, as links are changed from the Watts/Strogatz local neighborhood to random targets. On the other hand, the average path length also decreases remarkably.

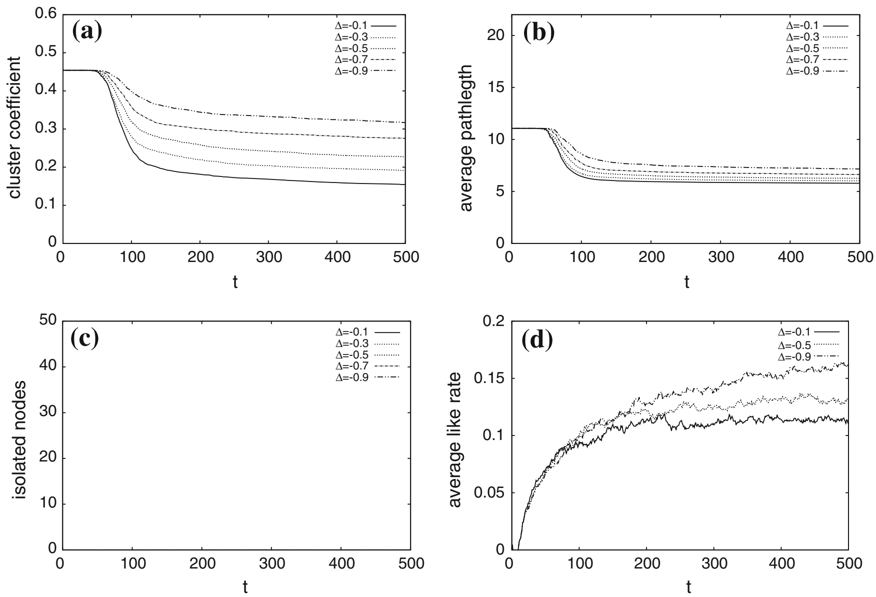


Fig. 2 Scenario 2—add random links after removal. **a** Clustering coefficient. **b** Average path length. **c** Isolated nodes. **d** Average like rate

Both changes are dependent on the threshold Δ, where higher values generally lead to lesser decreasing of the clustering coefficient but also only lesser positive impact on the path length (Fig. 2).

Although no more nodes are isolated, the user graph still regularly looses its connectivity over time without notable influence of the threshold Δ. This can also easily be explained by the underlying transformation to a random graph, which has no guarantee to be connected at all. The average like rate for Δ closer to 0 is even worse than it was in scenario 1 without adding new links at all. Again, the obvious explanation lies in the randomness of the newly added links, meaning that the newly added neighbor may as well be "worse" (in the sense of similarity to the user's interests) than the previously deleted link.

This approach is of course not realistic, since real-world users will neither simply select new contacts randomly and are not able to browse a whole OSN randomly to do so. Instead, new contacts will more likely be found in a user's local neighbourhood and only very seldomly in more distant areas of the network.

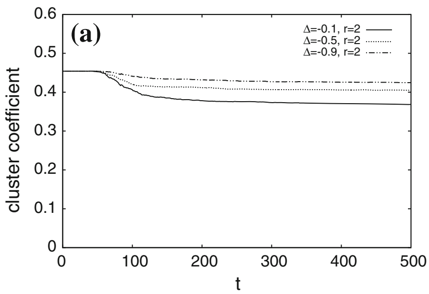
4.3 Scenario 3—Best Fitting Neighbors

In the third scenario, we therefore let a user try to replace removed link with a link to a user more similar to itself from within the local neighbourhood. When a link is removed, the affected users actively parse its neighborhood to find the neighbor with the highest similarity to the user's interest areas. The interest areas of all neighbors within r hops are compared to the interest area of the current user. The neighbor with the highest similarity is added to the local neighborhood of the user who initiated the search. Since all neighbours within 1 hop are already neighbors, r starts with a value of 2. In the simulations discussed in this article, the maximum value for r was chosen to be 4, which also seems to be a reasonable practical limit.

This approach still breaks up some local connections, so that the clustering coefficient still decreases in comparison to the initial Watts/Strogatz small world. Nevertheless, it remains much closer to the initial value than it did in the previous scenario. The influence of Δ is similar to the random scenario, where values close to 0 lead to stronger decrease of the clustering coefficient than values closer to -1.0 (see Fig. 3a). The impact increases for larger values of r, because in this case, new links can be added to more distant users, which will most likely not have any connections to the current user's neighbors and hence lead to a lower local clustering coefficient (Fig. 3b).

The influence on the average path length of both Δ and r is similar to the influence on the clustering coefficient (Fig. 4c, d). The closer Δ is to -1.0, the lesser the path length decreases. The higher the value of r is, the more the path length decreases. The average like rate increases in all cases (Fig. 4a, b), converging almost to the same value for all Δ and only to minor differences over different values of r.

Although loosing graph connectivity still remains as unsolved problem, the approach to add links by performing a local similarity search seems to be an appropriate solution for handling necessary removals of connections to other users.



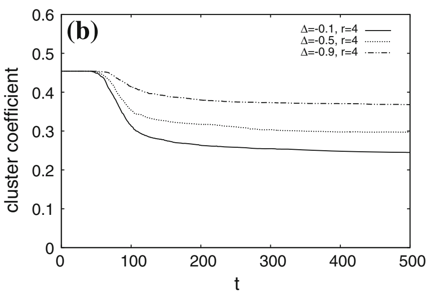


Fig. 3 Scenario 3—add best fitting neighbour, cluster coefficients. **a** Clustering coefficient $r = 2$. **b** Clustering coefficient $r = 4$

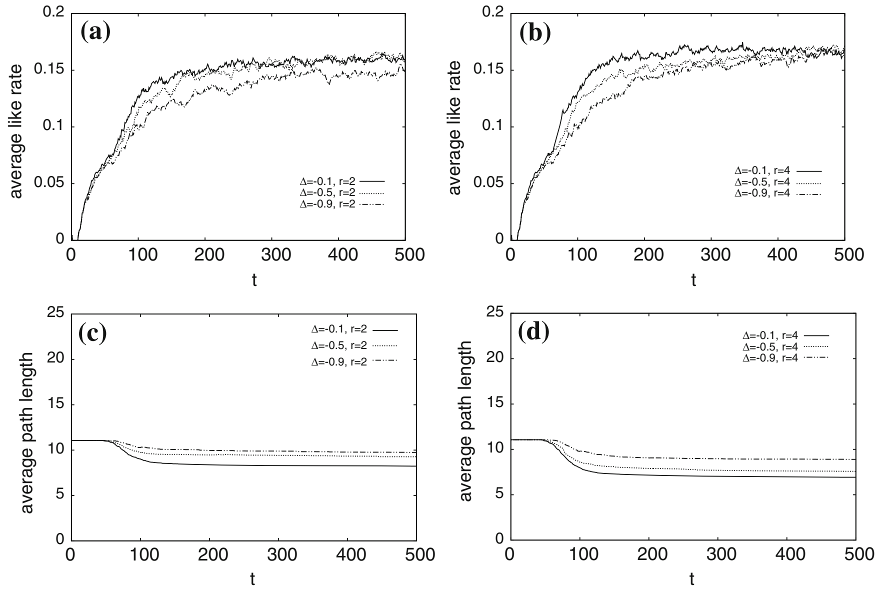


Fig. 4 Scenario 3—add best fitting neighbour, like rate and path length. **a** Average like rate $r = 2$. **b** Average like rate $r = 4$. **c** Average path length $r = 2$. **d** Average path length $r = 4$

5 Summary and Outlook

In this article, we have discussed an extension of our basic user centric model for content distribution in online social networks. It was shown that the introduction of link removal and link replacement based on the ongoing evaluation of recommendations from the user's perspective allows for an increasing average like-rate but negatively impacts the user graph's connectivity and local clustering. These effects can be moderated but not completely eliminated, when removed links are replaced using a best-fitting strategy in a user's local neighborhood. Future works will analyze the influence of these structural changes on the content distribution in the network, especially with regard to occurrence and distribution of superpopular contents.

References

1. Albert, R., Barabási, A.: Statistical mechanics of complex networks. Rev. Mod. Phys. **74**(1), 47–97 (2002)
2. Barabási, A., Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
3. Cha, M., Mislove, A., Gummadi, K.P.: A measurement-driven analysis of information propagation in the flickr social network, In: 18th International Conference on World Wide Web, pp. 721–730. ACM, New York (2009)

4. Coltzau, H., Unger, H.: A model based approach on superpopular contents in online social networks. In: 8th GI Conference on Autonomous Systems, pp. 145–156. VDI Verlag, Düsseldorf (2015)
5. Doerr, B., Fouz, M., Friedrich, T.: Why Rumors spread so quickly in social networks. CACM **55**(6), 70–75 (2012)
6. Easley, D., Kleinberg, J.: Networks, crowds, and markets: reasoning about a highly connected world. Cambridge University Press (2010)
7. Jamali, M., Haffari, G., Ester, M.: Modeling the temporal dynamics of social rating networks using bidirectional effects of social relations and rating patterns. In: 20th International Conference on World Wide Web, pp. 527–536. ACM, New York (2011)
8. Kleinberg, J.: The small-world phenomenon: an algorithmic perspective, In: Thirty-second annual ACM symposium on Theory of computing, pp. 163–170. ACM, New York (2000)
9. Lerman, K., Ghosh, R.: Information contagion: an empirical study of the spread of news on Digg and Twitter social networks. In: 4th International Conference on Weblogs and Social Media. AAAI (2010)
10. Schiöberg, D., Schmid, S., Schneider, F., Uhlig, S., Schiöberg, H., Feldmann, A.: Tracing the birth of an OSN: social graph and profile analysis in Google+. In: 4th Annual ACM Web Science Conference, pp. 265–274. ACM, New York (2012)
11. Toivonen, R., Onnela, J., Saramäki, J., Hyvönen, J., Kaski, K.: A model for social networks. In: Dawson, K.A., Indekeu, J.O., Stanley, H.E., Tsallis, C. (eds.) Physica A, vol. 371, no. 2, pp. 851–860. Elsevier (2006)
12. Viswanath, B., Mislove, A., Cha, M., Gummadi, K.P.: On the evolution of user interaction in Facebook, In: 2nd ACM Workshop on Online Social Networks, pp. 37–42. ACM, New York (2009)
13. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**, 440–442 (1998)
14. Zhang, M., Yang, P., Tian, C., Xiang, C., Xiong, Y.: Walk globally, act locally: efficient influential user identification in mobile social networks. In: First International Workshop on Mobile Sensing, Computing and Communication, pp. 29–34. ACM, New York (2014)

http://www.springer.com/978-3-319-40414-1

Recent Advances in Information and Communication Technology 2016
Proceedings of the 12th International Conference on Computing and Information Technology (IC2IT)
Meesad, P.; Boonkrong, S.; Unger, H. (Eds.)
2016, XIII, 334 p. 118 illus., Softcover
ISBN: 978-3-319-40414-1