Adaptive topology evolution in information-sharing social networks

Duanbing Chen,[1] Giulio Cimini,[2,*] Linyuan Lü,[2] Matúš Medo,[2] Yi-Cheng Zhang,[1,2] and Tao Zhou[1,2]

[1] *Web Sciences Center, School of Computer Science, University of Electronic Science and Technology of China, Chengdu 611731, P.R. China*

[2] *Physics Department, University of Fribourg - CH-1700 Fribourg, Switzerland*

(Dated: July 25, 2011)

The advent of Internet and World Wide Web has led to unprecedent growth of the information available. People usually face the information overload by following a limited number of sources which best fit their interests. In order to get the picture it is important to address issues like who people do follow and how they search for better information sources. In this work we conduct an empirical analysis on different on-line social networking sites, and draw inspiration from its results to present different source selection strategies in an adaptive model for social recommendation. We show that local search rules which enhance the typical topological features of real social communities give rise to network configurations that are globally optimal. Hence these abstract rules help to create networks which are both effective in information diffusion and people friendly.

PACS numbers: 89.65.Ef, 89.75.Fb, 89.75.Hc

I. INTRODUCTION

The fast development of the Internet has caused the amount of information available to grow dramatically. Therefore, people can hardly find what they are interested in. The problem of delivering the right content to the right person has attracted much attention in recent years. A possible solution is represented by Recommender Systems [1–3], which act as personalized information filters by analyzing users' profiles and past activities. Techniques used to produce recommendations include Collaborative Filtering [2, 4], Bayesian clustering [5], Probabilistic Latent Semantic Analysis [6], matrix decomposition [7, 8], diffusion and conduction [9–11] and many others. However it was recently shown that similarity of users' past activities plays a less important role than social influence: people value recommendations obtained by abstract mathematical analysis less than those coming from their friends or peers [12]. *Social recommendation* has hence emerged as a new approach which makes direct use of the social connections between members of a society [13]. Examples of social recommending implementations include services like Delicious.com, Flickr.com, LiveJournal.com, Youtube.com, FriendFeed.com and Twitter.com, where users can select some other users as information sources and *follow* them by importing or receiving respectively their URLs, photos, journals, videos, feeds and microblogs. In these systems the information spread from a user to her followers, and eventually to the followers' followers, and so forth. This diffusion mechanism resembles the spreading of epidemics or rumors over a network [14, 15].

A recently proposed news recommendation model [16–18] mimics the spreading process typical for social systems and combines it with an adaptive network of connections. In this model, when a user reads a news (or a different kind of content), she can either "approve" or "disapprove" it. If approved, the news spreads to the user's followers. Thus each user receives pieces of news from other users who represent her current *leaders* (i.e. information sources). Simultaneously with the spreading of news the leader-follower network evolves with time in order to connect users with similar tastes. A key aspect of this model is hence how to find good sources for each user. In [17] the authors propose a hybrid strategy for leaders updating based on local search and random off-trap, that is able to efficiently optimize the network of connections. The local aspect of the proposed strategy considers the leaders of her current leaders as potential candidates for each user, increasing in this way the *clustering coefficient* of the network. However this approach leaves aside other potential good candidates. For instance, real life examples reveal that a follower of a user is very likely to become a good leader for her too, as suggested by the high value of the *link reciprocity* in many information-sharing social networking services.

In this work we first conduct an empirical analysis on different on-line social networks, showing that real social communities are characterized by high values of link reciprocity and clustering coefficient. Then, building on the model introduced in [16], we propose and study different local leader updating strategies, and we compare the features of the resultant network topology from the viewpoint of user' satisfaction, network adaptation and recommendation efficiency. We only rely on local search rules because centralized-search mechanisms are very demanding and almost unfeasible for large-scale networks. Besides we wish to study the evolution of social and peer-to-peer networks, where users do not have a complete view of the network but only a limited and localized one. However we show that this apparent drawback can be overcome by an apt choice of these rules: local awareness of the network becomes almost as effective as global knowledge in producing optimal topologies. Moreover we find that an effective local updating strategy actu-

* giulio.cimini@unifr.ch

ally enhances both reciprocity and clustering coefficient of the network, mimicking in this way the users' search of sources (or in general acquaintances) in social networks.

II. EMPIRICAL ANALYSIS

In this section we extract the features of five different on-line information-sharing social networking sites: *delicious.com*, *flickr.com*, *livejournal.com*, *youtube.com* and *friendfeed.com*. In these systems users form a social network and can share different kind of content—respectively bookmarks, photos, blog articles, videos and feeds. Table I gives an overview of the features of the different systems. Note that we excluded from the analysis both isolated nodes and self-loops. To describe the networks' topologies we use two standard quantities:

- *Link reciprocity* (r) is the tendency of node pairs to form connections between each other and is defined as the ratio of the number of bi-directed links to the total number of links in the network [19].
- *Clustering coefficient* (c) measures the tendency of the network to form tightly connected components and is defined as the ratio of the number of directed link triangles that exist among a user and her first neighbors to the total number of triangles that can exist among these users, averaged over all users [20]:

Delicious.com, previously known as *del.icio.us*, is the world-largest online bookmarking website. Users in delicious.com collect URLs as bookmarks; moreover, they can select other users to be their *leaders* (i.e. information sources) and *follow* them by importing their own bookmarks. Hence we can naturally represent the delicious.com community by a directed leader-follower network. To extract the network's structure, we perform a crawl of the users graph by accessing the public web interface provided by the site: starting from a user, we follow her outgoing and incoming links to reach other users, and so on. This algorithm is known as breadth-first search (BFS) [21]. The dataset is being collected since May 2008, and it consists of 854,357 users and 2,521,187 directed links among them; out of these users, more than 99% belong to the giant component. The values of the reciprocity and the clustering coefficient for the delicious.com network are reported in Table I.

We also report the empirical results for *flickr.com*, *livejournal.com* and *youtube.com*. In these website users can select other users as friends (leaders, as intended in this paper) to get access to their content (photos, blogs and video respectively). The leader-follower networks of these on-line communities were obtained in [22] by crawling the large weakly connected component of the corresponding user graphs. The algorithm used for the crawl was again BFS with snowball method [23]: the data extraction starts from a set of seed users and then it expands by following the outgoing links of these users to reach new users, and so on.

Friendfeed.com is a microblogging service created in 2007 and acquired by Facebook in 2009, in which users can share short messages to a list of contacts, who can comment back under the original messages. It is also a feed aggregator, importing data from several other services like Twitter, Facebook, YouTube, Flickr and Google Reader. The leader-follower network we analyze was crawled in [24].

The summary of the results is reported in Table I. We immediately notice that both the level of link reciprocity and the degree of local clustering in all social networks are significantly high. For comparison, we also report the values of the reciprocity r_0 and clustering coefficient c_0 of Erdös and Rényi random graphs [25] with the same number of nodes and links as the real networks. Both r_0 and c_0 are given by the ratio of the actual number of links to the total number of possible links in the graphs, hence $r_0 = c_0$. These values turn out to be between two and five orders of magnitude smaller than what we observe in real social networks. This phenomenon has a natural explanation in information-sharing social communities: if two users have common interests each of them can likely provide the other with the right content; also, people tend to be introduced to other people via mutual friends, increasing the probability that two friends of a single user are also friends. In the following sections we will draw inspiration from these observations to define the topology evolution rules of an adaptive model for social recommendation.

III. MODEL DESCRIPTION

We now briefly summarize the news recommendation model based on [16, 17] that will be used for the study of different leader selection strategies.

The system consists of U users, each connected to L other users (the user's leaders) by directed links. The value of L is fixed as users usually follow a limited number of sources. Users receive news from their leaders and eventually read and rate them; in addition, they can introduce new content to the system.

Evaluation of news α by user i ($e_{i\alpha}$) is either $+1$ (liked), -1 (disliked) or 0 (not read yet). Similarity of reading tastes of users i and j (s_{ij}) is estimated by comparing past users' assessments: if i and j evaluated N_{ij} news in common and agreed A_{ij} times, their similarity can be measured in terms of the overall probability of agreement

$$s_{ij} = \frac{A_{ij}}{N_{ij}} \left(1 - \frac{1}{\sqrt{N_{ij}}} \right) \tag{1}$$

where the term in the parentheses disadvantages user pairs with small overlap N_{ij} (which are more sensitive to statistical fluctuations). For $N_{ij} \leq 1$, s_{ij} is replaced by a small positive value s_0. Apart from their ratings, no other information about users is assumed by the model.

Propagation of news works as follows. When news α is introduced to the system by user i at time t_α, it is

TABLE I. Statistics of social networking sites.

	Delicious	Flickr	LiveJournal	YouTube	Friend-Feed
Date of crawl	05-2008	01-2007	12-2006	01-2007	09-2009
Number of users	854,357	1,715,255	5,203,764	1,138,499	513,588
Number of links	2,521,187	22,613,980	76,937,805	4,945,382	19,810,789
Reciprocity	0.392	0.624	0.734	0.791	0.207
Clustering	0.161	0.165	0.255	0.077	0.146
r_0, c_0	$3.45{\cdot}10^{-6}$	$7.68{\cdot}10^{-6}$	$2.84{\cdot}10^{-6}$	$3.81{\cdot}10^{-6}$	$7.51{\cdot}10^{-5}$

passed from i to her followers j with a *recommendation score* proportional to their similarity s_{ij}. If this news is later liked by one of users j who received it, it is similarly passed further to this user's followers k (with recommendation score proportional to s_{jk}), and so on. For a generic user k at time t, news α is recommended according to its current recommendation score:

$$R_{k\alpha}(t) = \delta_{e_{k\alpha},0}\, \lambda^{t-t_\alpha} \sum_{l\in L_k} s_{kl}\, \delta_{e_{l\alpha},1} \qquad (2)$$

Here L_k is the set of leaders of user k, the term $\delta_{e_{k\alpha},0}$ equals one only when user k has not read news α yet and the term $\delta_{e_{l\alpha},1}$ is one only if user l liked news α. The sum represents the instance of a user receiving the same news from multiple leaders—recommendation scores are summed up in this case, reflecting that a news liked by several leaders is more likely to be liked by this user too. Finally, to allow fresh news to be accessed fast, recommendation scores are exponentially damped with time, with $\lambda \in (0, 1]$ being the damping factor.

Starting from an initial random network configuration (random assignment of leaders to users), connections are periodically rewired to drive the system to an optimal state where users with high similarity (taste mates) are directly connected. When rewiring occurs for user i, the leader with the lowest similarity value (j) is replaced with a new user (k) if $s_{ik} > s_{ij}$. There are different selection strategies for picking new candidate leaders:

1. *Random rewiring.* k is simply a user picked at random in the network.

2. *Local rewiring.* k is the user in the neighborhood of user i with the maximum value of s_{ik}. This mechanism is based on the observation that two users who have common acquaintances are likely to have similar interests. As will be discussed in the next section, there are different ways to define such neighborhood.

3. *Hybrid rewiring.* Random rewiring is used in some cases and local rewiring in the others. This mechanism mimics both users searching for friends among friends of friends (local rewiring) and having casual encounters which may lead to long-term relationships (random rewiring).

4. *Global rewiring.* k is the user who maximizes s_{ik} among all users U (this is a local rewiring with the neighborhood being the whole network).

A. Topology evolution

The search for new and better information sources is a fundamental feature of many social communities. In the model described above, the leader updating procedure is intended to drive the network to an optimal state where users with high similarity are directly connected, so that the system is able to efficiently deliver right news to right users. We remind here that one is constrained to local search rules because global search mechanism (such as the *Global rewiring*) are very demanding for large-scale networks and also unfeasible without a centralized control. On the other hand, the *Random rewiring* strategy is very inefficient as good new leaders are hardly found by chance. Besides we want to be true to life: users always have a little and localized view of the network. Therefore we shall define the "neighborhood" of a user, i.e. a set of close users in the network who stand for possible candidate leaders. This is the basis of the *Local rewiring.* The choice of a specific neighborhood should be clever enough to allow users to actually find their taste mates. For instance, the pool of candidate leaders should not be too large, as in this case the search becomes unmanageable both from the system's and the users' point of view. On the other hand, if the neighborhood size is very small (compared to the whole network), the rewiring may stop at a sub-optimal assignment of leaders: the topology evolution halts if users' better leaders are at some moment out of the neighborhoods (they can never be reached), meaning that the algorithm got trapped in a sub-optimal state [17]. A possible solution to this problem is to employ some percentage of randomness in the selection, as in the *Hybrid rewiring.* In this way users may happen to get connected regardless of their distance, and the pool of candidate leaders for each user is potentially the whole network. In the following analysis we will always make use of a Hybrid rewiring with 10% of randomness, to exploit mainly the local search but to avoid trapping in a local minimum (see [17] for a detailed study of the effect of the randomness percentage on the rewiring efficiency).

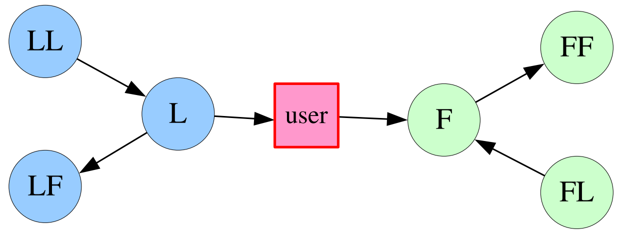

FIG. 1. Local network structure of one user. Links' directions reflect how information flows between users.

B. Neighborhood definition

We shall now define the "neighborhood", i.e. the set of candidate leaders exploited by the Hybrid rewiring. The local network structure from a user's viewpoint is represented in Figure 1. At distance one from the user there are two sets of users: her *leaders* (L) and *followers* (F). L and F form the *first layer* from the user. At distance two, we find four different sets of users: her *leaders' leaders* (LL), *leaders' followers* (LF), *followers' leaders* (FL) and *followers' followers* (FF). These sets form the *second layer* from the user. Notice that the described sets may overlap with each other (e.g. a user can be leading but also following another user). Given such scheme of the local network structure, we have to consider which of these sets contain potential good leaders for the user.

Apart from the current set of leaders L, the first layer contains a good set of candidates—represented by F. Indeed if user i is a good leader of user j, meaning that j obtains valuable information from i, then i and j are likely to have some common interests and the similarity between them can be high. Hence also user j can provide user i with the right content and be a good information source for her. This assumption is supported by the high value of the *link reciprocity* in many information-sharing social networks (see Table I). Including F in the candidate set hence increases the probability of having reciprocal links. However, this set may be too small to be considered alone. Therefore we move further to the second layer. The leaders' leaders set (LL) was considered in [17] where the authors observed that since user j obtains valuable information from her leader i and such information may come from i's leaders, then j can have similar interests with i's leaders and benefit from following them. Again this assumption is supported by the high value of the *clustering coefficient* in many social networks (Table I). Analogous considerations lead us to take into account also the LF, FL and FF sets.

In the following sections we will study the behavior of the described model for different definition of the neighborhood. When using Hybrid rewiring, we will simply denote it by the neighborhood that it exploits. For instance it will be named as LL if only leaders' leaders are considered as candidates, and $LL+F$ if also followers are included.

IV. RESULTS

For numerical tests of the model, we use an agent-based framework. Tastes of user i are represented by a D-dimensional binary vector $\mathbf{t}_i$ and attributes of news α by a D-dimensional binary vector $\mathbf{a}_\alpha$. Each vector has a fixed number, D_A, of elements equal one (active tastes) and all remaining elements equal zero. We always set the system so that all mutually different user taste vectors are present exactly once: $U = \binom{D}{D_A}$. This also means that the taste vectors of two users differ at least in two elements. Hence we define as "taste-mates" users with exactly two different taste vector elements. Opinion of user i about news α is based on the overlap of the user's taste vector with the news's attribute vector

$$\Omega_{i\alpha} = (\mathbf{t}_i, \mathbf{a}_\alpha) \tag{3}$$

where $(\cdot,\cdot)$ is a scalar product of two vectors. If $\Omega_{i\alpha} \geq \Delta$ user i likes news α ($e_{i\alpha} = +1$), otherwise she dislikes it ($e_{i\alpha} = -1$). Here Δ is the users' approval threshold.

Simulation runs in discrete time steps. In each step, an individual user is active with probability p_A. When active, the user reads and evaluates the top R news from her recommendation list and with probability p_S submits a new news with attributes identical to the user's tastes. The network of connections is rewired every u time steps.

Parameters values used in all following simulations are: $D = 14$, $D_A = 6$ (so that $U = 3003$), $L = 10$, $p_A = 0.05$, $p_S = 0.02$, $R = 3$, $\Delta = 3$, $\lambda = 0.9$, $s_0 = 0.001$, $u = 10$. To measure the system's performance we use:

- *approval fraction*, the ratio of news approvals to all assessments: it tells us how often users are satisfied with the news they read.
- *average differences*, the average number of vector elements in which users differ from their leaders: it measures how well the network has adapted to users' tastes.

Figure 2 shows the approval fraction (a) and the average differences (b) at different times steps of the network's evolution and for different definitions of the neighborhood exploited by the Hybrid rewiring. Global and Random methods are shown as benchmarks. As expected, if we limit the pool of candidate leaders to F, users are not much satisfied because they can hardly find good information sources. This is the result of having considered a very small set (the average number of followers for a user equals L). If instead we define LL as the neighborhood (as in [17]), we significantly improve both users' satisfaction degree and network's adaptation speed. This is because the candidate set is much wider in this case—there are on average $L[L-r-(L-1)(c+Lq/2)]$ different leaders' leaders for a users, and this number is much greater than L for typical values of r, c and q (here q is the probability that four users are linked in a closed square structure). To further improve the performance of the system, we expand the candidate pool to $LL+F$.

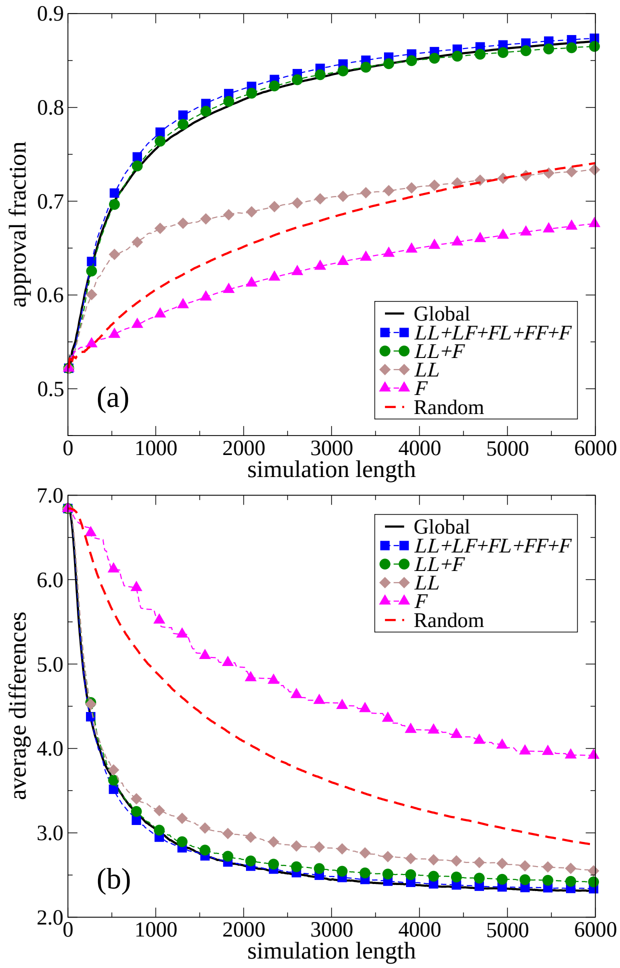

FIG. 2. Approval fraction (a) and average differences (b) for different rewiring strategies. The lowest value of the average differences is two, as it is the taste vector differences between taste-mates.

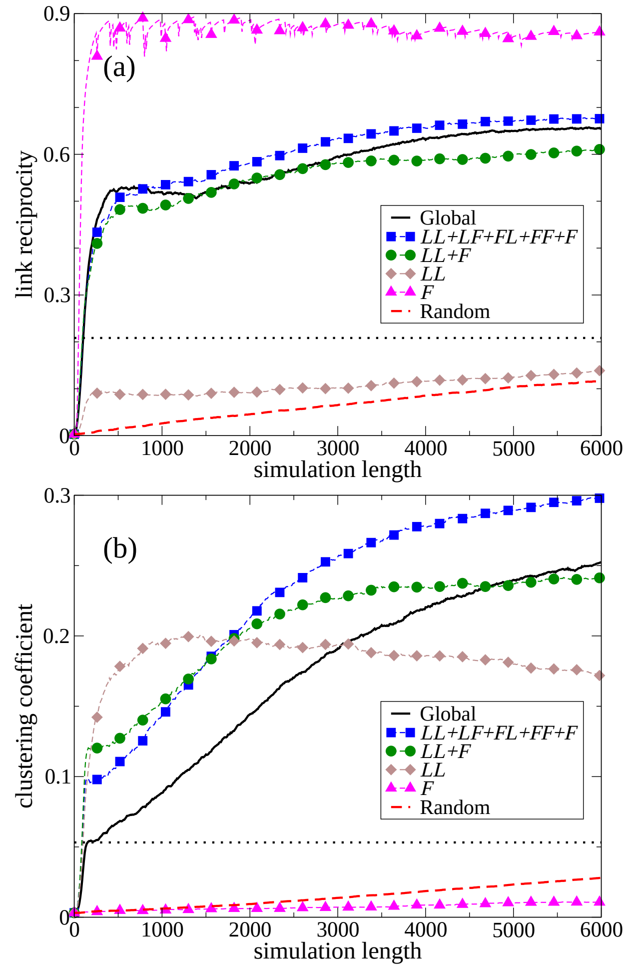

FIG. 3. Link reciprocity (a) and Clustering coefficient (b) for different rewiring strategies. The values of r_* and c_* are each represented by a horizontal dotted line in the respective plots.

With this definition of the neighborhood we promote at the same time the reciprocity and the clustering coefficient of the network, obtaining a surprising effect: both approval fraction and average differences become as good as the ones obtained by the Global rewiring, i.e. by considering the whole network as the candidate leader set. In other words, such a small local scale turns out to be as representative as a whole-network scale. Hence further expanding the candidate set to the whole second layer ($LL + LF + FL + FF + F$) does not bring to any substantial improvements. We remark that this feature does not depend on the size of the system. We run simulations of a 15-times bigger system with $U = \binom{18}{8} = 43758$ and $\Delta = 4$, and observe that the simple $LL + F$ has again the same performance as the global search.

We also measure the values of link reciprocity and clustering coefficient in the network. The evolution of these quantities is shown in Figure 3. We firstly introduce two reference values for r and c. In the initial random network the average probability to find a reciprocal link between two connected vertices is simply equal to the average probability of finding a link between any two vertices, which is given by $(UL)/[U(U-1)] = L/(U-1)$. Hence we have $r_0 := r(t = 0) = L/(U-1)$. This statement also holds for the probability to find a closed link triangle between three users, i.e. for the clustering coefficient: $c_0 := c(t = 0) = L/(U-1)$. Instead if the network is in a structure-less optimal configuration where each user has randomly chosen L of her $N = D_A(D - D_A)$ taste-mates as leaders, then the value of the reciprocity becomes $r_* = L/N$. Besides in this network state the probability that two taste-mate neighbors of a user are also taste-mates with each other is given by $(D-2)/(N-1)$. To show this, consider two taste mate users: there are $(D_A - 1) + (D - D_A - 1) = D - 2$ other users who are taste mates with both of them and $(D_A - 1)(D - D_A - 1) = N - D + 1$ who are taste mates with only one of them. The clustering coefficient is given by the above-mentioned probability conditional to the existence of a link: $c_* = [(D-2)/(N-1)][L/N] = [L(D-2)]/[N(N-1)]$. Figure 3a shows that the reci-

TABLE II. Specificity $(1-\alpha)$ and Sensitivity $(1-\beta)$ of recommendation for different rewiring strategies.

	$1-\alpha$	$1-\beta$
Global	95.9%	12.6%
$LL+FL+FL+FF+F$	96.3%	10.9%
$LL+F$	96.6%	10.6%
LL	94.8%	9.2%
F	81.8%	29.9%
Random	84.2%	33.1%

procity coefficient grows from r_0 as the system evolves with any rewiring method. As expected, the value of r is very high with F (by construction, F promotes reciprocity), and low for Random and LL. In the latter case, r will eventually converge to r_*. The other methods achieve similar values of r, which are comparable with the reciprocity degree of real social networks (Table I). The clustering coefficient (Figure 3b) shows an opposite trend: it becomes soon very large with LL (by construction) while it remains quite small for Random and F, eventually converging to c_*. For the other methods c converges to values again comparable with the clustering coefficient of real social networks (Table I). The differences between the various methods are more evident for c than for r, as including more sets from the second layer will result in increasing the probability to form a closed link triangle

At last, we discuss the efficiency of the modeled recommender system. When making recommendations, it is possible to fall into two different kinds of error: recommending content that users wouldn't like, and not recommending content that users would like. These errors are known respectively as of type I (*false positives*) and of type II (*false negatives*) [26]. To complete the picture, *true positives* are recommendations of content that users would like, and *true negatives* are lacks of recommendation of content that user wouldn't like. Note that false positives upset users but false negatives do not (i.e. a type I error has more serious consequences than the other), hence a good recommendation engine should mainly reduce false positives. We further introduce the *specificity* $(1-\alpha)$ and the *sensitivity* $(1-\beta)$ of the recommendation system as the ability to avoid respectively false positives and false negatives:

$$1-\alpha=\frac{\text{TN}}{\text{TN+FP}} \quad , \quad 1-\beta=\frac{\text{TP}}{\text{TP+FN}}$$

where TP, TN, FP and FN are respectively the number of true positives, true negatives, false positives and false negatives. To measure these quantities in our artificial setting, we define α as the average number of wrong recommendations for a news over the number of users who might dislike this news—given by $\sum_{k=0}^{\Delta-1}\binom{D_A}{k}\binom{D-D_A}{D_A-k}$, and $1-\beta$ as the average number of good recommendations for a news over the number of users who might like this news—given by $\sum_{k=\Delta}^{D_A}\binom{D_A}{k}\binom{D-D_A}{D_A-k}$. Table II reports the stationary values of specificity and sensitivity for the recommender system when different source selection strategies are employed. Specificity is remarkably high for all methods, especially for the best performing ones, hence the number of false positives in the system is very low. Sensitivity shows instead an opposite trend: Random and F updating strategies are the best performing now. We see that the effort of reducing one type of error results in increasing the other type, as it generally happens in statistical tests. In our case the reason behind this phenomenon is the presence of tightly connected components in the system: in a highly clustered network news have few paths to spread far from the users who post them (and the spreading process takes long time), hence they tend to remain localized. As a consequence, few users receive a news but almost all of them like it. When clustering is low, a news has more spreading directions, hence it can reach many users but more of them eventually dislike it. However we are mainly interested in having a recommender system with high specificity, and in this sense simple local strategies (like $LL+F$) again perform at the same level of global search in generating optimal network structures for recommending and sharing information.

V. CONCLUSION

How to recommend the right content to the right person and what/who are this person's favorite information sources are fundamental questions in the age of information overload. In this work we exploited a recently proposed news recommendation model which combines similarity of users' past activities and social relationships to obtain recommendations, and which mimics the spreading process typical for social systems where the network of connections continually evolves with time [16]. The topology evolution serves users looking for newer and better information sources. Since global optimization of the users' connections is computationally prohibitive for a large system, a key issue of the model is where to find good new leaders for users. Taking real life as inspiration, we try to model the users' search of sources in real social communities, retaining that users only have a limited view of the network. For this purpose we designed different local search strategies which increase the network's reciprocity and clustering coefficient. We then studied the resulting evolution and properties of the system and showed that with these local search rules the users' community can self-organize into optimal topologies, almost equivalent to the ones that can be generated by global knowledge of the system. Indeed the resulting artificial networks have high reciprocity and clustering, as similar to the real information-sharing social communities studied in section II. Therefore our automated abstract rules help to create networks which not only are effective for the spreading of information but also resemble structures resulting from real human activity.

ACKNOWLEDGMENTS

This work was partially supported by the Future and Emerging Technologies programmes of the European Commission FP7-ICT-2007 (project LiquidPublication, grant no. 213360) and FP7-COSI-ICT (project QLectives, grant no. 231200), by the Swiss National Science Foundation (grant no. 200020-121848), by the National Natural Science Foundation of China (grant nos. 60973069, 90924011 and 60903073) and by the International Scientific Cooperation and Communication Project of Sichuan Province in China (grant no. 2010HH0002).

[1] P. Resnick and H. R. Varian, Comm. ACM **40**, 56 (1997)
[2] J. L. Herlocker, J. A. Konstan, L. G. Terveen and J. T. Riedl, ACM Trans. Inf. Syst. **22**, 5 (2004)
[3] G. Adomavicius and A. Tuzhilin, IEEE Transactions on Knowledge and Data Engineering **17**, 734 (2005)
[4] G. Linden, B. Smith and J. York, IEEE Internet Computing **7**, 76 (2003)
[5] J. S. Breese, D. Heckerman and C. Kadie, in *Proc. 14th Conference on Uncertainty in Artificial Intelligence* (1998)
[6] T. Hofmann, ACM Trans. Inf. Syst. **22**, 89 (2004)
[7] S. Maslov and Y.-C. Zhang, Phys. Rev. Lett. **87**, 248701 (2001)
[8] Y. Koren, R. Bell and C. Volinsky, Computer **42**, 30 (2009)
[9] T. Zhou, J. Ren, M. Medo and Y.-C. Zhang, Phys. Rev. E **76** 046115 (2007)
[10] Y.-C. Zhang, M. Blattner and Y.-K. Yu, Phys. Rev. Lett. **99** 154301 (2007)
[11] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. R. Wakeling and Y.-C. Zhang, Proc. Natl. Acad. Sci. USA **107** 4511 (2010)
[12] R. Sinha and K. Swearingen, in *Proc. of the DELOS-NSF Workshop:Personalisation and Recommender Systems in Digital Libraries* (2001)
[13] J. Golbeck, Science **321**, 5896 (2008)
[14] T. Zhou, Z. Fu and B.-H. Wang, Prog. Nat. Sci. **16** 452 (2006)
[15] Y. Moreno, M. Nekovee and A. F. Pacheco, Phys. Rev. E **69** 066130 (2004)
[16] M. Medo, Y.-C. Zhang and T. Zhou, Eur. Phys. Lett. **88** 38005 (2009)
[17] G. Cimini, M. Medo, T. Zhou, D. Wei and Y.-C. Zhang, Eur. Phys. J. B **80** 201 (2011)
[18] D. Wei, T. Zhou, G. Cimini, P. Wu, W. Liu and Y.-C. Zhang, Physica A **390** 2117 (2011)
[19] S. Wasserman and K. Faust, *Social Network Analysis*, Cambridge University Press, Cambridge (1994)
[20] G. Fagiolo, Phys. Rev. E **76** 026107 (2007)
[21] C. Y. Lee, IRE Transactions on Electronics Computers EC-10, **2** 346 (1961)
[22] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel and B. Bhattacharjee, in *Proceedings of the 5th ACM/USENIX Internet Measurement Conference* (2007)
[23] S. H. Lee, P. J. Kim and H. Jeong, Phys. Rev. E **73** 016102 (2006)
[24] F. Celli, F. M. L. Di Lascio, M. Magnani, B. Pacelli and L. Rossi, in *International Conference on Social Computing, Behavioral Modeling and Prediction* (2010)
[25] P. Erdős and A. Rényi, Publicationes Mathematicae Debrecen **5** 290 (1959)
[26] J. Neyman and E. S. Pearson, Biometrika **20A** 175 (1928)