Diversity in open social networks

Daniel Lemire, Stephen Downes, Sébastien Paquet

Abstract

Online communities have become become a crucial ingredient of e-business. Supporting open social networks builds strong brands and provides lasting value to the consumer. One function of the community is to recommend new products and services. Open social networks tend to be resilient, adaptive, and broad, but simplistic recommender systems can be 'gamed' by members seeking to promote certain products or services. We argue that the gaming is not the failure of the open social network, but rather of the function used by the recommender. To increase the quality and resilience of recommender systems, and provide the user with genuine and novel discoveries, we have to foster diversity, instead of closing down the social networks. Fortunately, software increases the broadcast capacity of each individual, making dense open social networks possible. Numerically, we show that dense social networks encourage diversity. In business terms, dense social networks support a *long tail*.

Keywords: social networks; recommender systems; long tail

1. Introduction

Social networks are sets of relationships created by interactions among individuals. They constitute the fundamental ingredient of many innovation-diffusion models [12,13] as well as viral-marketing models [14]. We distinguish two types of social networks. Closed networks have barriers to entry and exit from the network, and connections between individuals may be limited. In open networks, there are no such barriers. In particular, any two individual may decide to form an ad hoc relationship. Private clubs, companies, tribes, countries, industry organizations and professional corporations are examples of closed social networks. Although restricted, closed social networks can be democratic or even operate by consensus. By contrast, while in open social networks, any user can join or leave the network at any time, the network itself might not be democratic or egalitarian. Similarly, while participation on the Web is open, there is no mechanism to ensure an equal visibility to all. The consumers of most large retailers form an open social network and so does the blogosphere, even though some members (A-listers) may be granted privileges or greater visibility. Being 'open' or 'closed' is a question of membership, not of governance.

One of the primary function of a social network is to guide its members. Recommender systems have become fashionable for their contribution to the success of some online

---

Published online on October 3rd 2008.

retailers. Back in 2006, Marshall stated that 35% of all sales at Amazon.com are due to their recommender system [10]. However, recommender systems have always existed: it seems to be part of human nature to recommend to others good locations or resources. As social networks become more open, more varied, and larger, their function as a recommender system can either improve or collapse. Even though contributors to open networks are often motivated by altruistic goals [39,40], online social networks have been 'gamed' by members seeking to profit from the network. For example, a popular Web-based recommender system, Digg, had to change its recommendation process since several users would collude to game the recommendations. Due to the popularity of Digg, these users may have had financial incentives to artificially boost the popularity of some news items. On this basis, Karp argued that completely open social networks have failed [9]. Should we close social networks to prevent the abuse of their recommender function? The ancestor of Wikipedia, Nupedia, was a closed peer system [38] and it failed while the open Wikipedia thrived: clearly openness and low barriers to entry can have tremendous benefits.

But what constitutes a good recommendation? The item being recommended has to have some intrinsic qualities, such as correctness and elegance, but it should also have qualities related to the community and the context, such as timeliness. Thus, the *intrinsic* qualities of an item are defined against a social network. For example, a scientific journal, through its review process, acts as a recommender system for scientists. If the editor of a journal sends an alchemy research paper to 100 alchemists, he may get good reviews. If he sends the same paper to a single chemist, he can expect a bad review. While traditional peer review relies on a closed social network, there is an ongoing debate as to whether it is the ideal system [11]. Alternative forms of peer review have been suggested, including posting papers online and letting the community decide what is useful: such an alternative would be based on an open network.

Users freely participate in open social network and do so only because the network is useful to them or their peers. Individual can break relationships when they are no longer useful. Thus, a good recommendation must be a natural output of an open social network, as users would simply leave the network otherwise. However, any specific processing, including peer review or commercial marketing, might be flawed even if it is supported by the network for a time.

Electronic media support larger open social networks. Through email or Web sites, an individual can interact with dozens, or even hundreds, of people a day. Computers can also automate the construction of aggregations to summarize the data collected in the social network. Digg, Slashdot, and many of Amazon's recommendation features are examples of aggregations. Aggregations do not define social networks, but they can play a pivotal role and they can be seen as the generalization of the simple recommendation. Software has the potential to increase the usefulness of open social network by increasing the density of the network: technology makes it possible for individuals to interact with more individuals. The mail and the phone made it possible for people to interact with friends far away, but the Internet has reduced dramatically the cost of each such interactions.

The failure of an aggregation to be resilient or useful does not signify that the social network has failed: the individuals composing the network will tend to stop using defective tools and may find or compose new ones. The recommender creates a view of the network, and it is this view which leaves the network open to gaming. For any large scale open social network, such a view is necessarily a simplified or reduced version of the social network: 1) it may only consider only part of the data and the connections 2) it aggregates several connections together for conciseness 3) it may be static, providing only a snapshot. In some cases, it may not even be possible for the software to consider all of the possible connections: the size of the World Wide Web can only be approximated [23,24]. Any attempt to present to a user a digestible summary requires making assumptions about the network and the users. This limited view of the network can be exploited.

Usefulness can be measured by the usage patterns of the recommendations. Because the recommendation process itself belongs to the social network, an analysis based on collected historical data maybe flawed unless we know how to model the users and their interactions with each other and the software. Nevertheless, we should seek guiding principles to develop more useful recommender systems.

2. What is diversity?

Diversity has long been valued in statistical surveying and research. Polling methodology, for example, requires that surveyors construct samples in such a way as to ensure a genuine cross-section of the population being polled; researchers attempt to match samples of the population to the population as a whole based on samples of various attributes observed in the population [47]. Coverage biases are a constant source of worry.

Diversity is defined as a high level of heterogenity in a collection of entities, that is, if the entities in the collection are *different* from each other. The difference between entities may be measured as the inverse of their *similarity*, as defined by Tversky (1977). In particular, Tversky writes, "the similarity of objects is expressed as a linear combination, or a contrast, of the measures of their common and distinctive features." The degree to which these features overlap is the degree to which two entities are similar, and consequently, the degree to which entities possess unique features is the degree to which they are different. A collection of entities is *diverse*, therefore, if the entities in that collection display a high degree of difference from each other, that is, if there is a high proportion of unique features (that is, features possessed by fewer, not more, entities) in the set.

Tversky's measure is similar to Shannon's information content [27] measure: suppose that in a set S of size n, the element i occurs with frequency f_i, then the information content is $-\sum f_i \log_2 f_i/n$. The monotonicity of information content follows from Shannon's seminal work. Using this last definition of diversity, given a fixed cardinality, the diversity is maximized when all items are distinct. In Shannon's system, an element plays the role of a feature in Tversky's measure. We might say that Tversky identifies the similarity of *entities*, while Shannon identifies the similarity of *messages*. Each type of similarity - and consequent diversity - plays a role in a network. Our work is directed toward the diversity of entities, but we suggest that similar considerations may apply regarding the diversity of messages.

We all know intuitively what a diverse set of entities is. Adapting an example from Tversky [25], Russia and Jamaica form a diverse set of countries compared to either Cuba and Russia, or Cuba and Jamaica. While there are several possible measures of diversity, we postulate that diversity measures over sets should obey the following rules or axioms:

- The diversity of a singleton or empty set is naught. A recommendation containing a single book has no diversity.
- Diversity is monotonic. Given sets A and B, the diversity of the union of the sets A and B is at least as large as the diversities of sets A and B alone. This implies that if A is a subset of B, then A has at most a diversity as large as B. That is, samples always have diversity no greater than the original set.
- A diversity measure is a non-negative number.

These rules do not imply that diversity is distributive: the diversity of the union of A and B may not be computable from the diversity of sets A and B alone. If we accept these axioms, then the set Cuba, Jamaica and Russia, has at least the diversity of the set Russia and Jamaica. A naïve measure of diversity is the cardinality. Using this measure, a larger set is automatically more diverse.

We can measure diversity geometrically. For example, if each data point can be represented as a feature vector, the number of dimensions spanned by the data points can be a measure of diversity. For example, if the feature vector is (in the Antilles, has an history of communism), then Russia = (0,1), Cuba = (1,1), Jamaica = (1,0) and the set Russia and Jamaica spans a two-dimensional space whereas Cuba and Russia spans a unidimensional space. The number of dimensions could also be a fractional number [26]. Dimensionality is always monotonic. Some structures may evolve over time. Even though any given snapshot of the structure may have low diversity, the overall structure, over time, may demonstrate diversity. For example, a conference behind held yearly, first in Russia, then in Cuba, then in Jamaica, would have chosen a diverse set of locations.

Thus, our axioms of diversity can be applied to graphs. For example, the maximal number of incoming links to a node is such a possible diversity measures. Many indicators related to diversity have been proposed on networks. The in-degree is the number of links to a node, the out-degree is the number of links out of a node, whereas the degree of a node is to the total number of links to other nodes from this node. Hubs in a network are nodes with high degree. We say that a network is disassortative if nodes having few degrees link to nodes having higher degrees. The betweenness is a measure of how frequently a given link is part of the shortest path between any two nodes: in a more diverse network no link should dominate all others and the betweenness should therefore tend to be flat. The average shortest distance between nodes is another related metric and so is the diameter of the graph (the largest distance between any two nodes).

Calculations of diversity are subject to perspective and point of view. While a theoretical diversity metric could be calculated with referenced to all features possessed by all entities in the set, such a measure is rarely possible or desired. As Tversky notes, calculations of similarity will take into account which of those features are *salient* or important to the calculation taking place. For people taking a tropical vacation, for4 example, Jamaica and Cuba are more similar, while for those interested in Cold War era governments, Cuba and Russia are more similar.

3. Diversity in Open Social Networks

In an open social network, the entities connected are individual people. In this environment, 'diversity' is typically defined as the distribution of a sample across a selection of representative population groups as identified by salient features, such as age, race, gender, language, location, and related properties (it could be argued that the salience of the features in such a set is related to the diversity of the messages they would send to each other).

In an open social network, there is naturally a constant flow of new individuals and changes in the structure of the network. So all things being equal, a social network that is open will also be diverse. But openness does not *entail* diversity; an open social network in a homogenous society would itself most likely be homogeneous. Also, self-selection may ensure some form of homogeneity. Moreover, while, diversity does not inherently require a large network, we expect worldwide open social networks to be diverse.

Given a large enough random sample, current polling methodologies predict accurately the result of democratic consultations. For other applications, Bourdieu [15] commented that we could not weight each individual as a unit and ultimately put into question the summary of a population by simple aggregations. John Stuart Mill and Alexis de Tocqueville referred to the tyranny of the majority: left unchecked, majorities may put their interests far above the minorities' interests. For this reason, most democratic systems do not rely exclusively on proportional votes. Ensuring that elected officials will come from various regions of a country using representative democracy is one approach to ensure some diversity.

Representativity in a social network can be obtained with democracy or random and fair samples. However, the representativity of the view is not essential nor necessarily useful. As an example, suppose that you have a probability p of appreciating any given recommendation. Suppose the recommendations are statistically independent. If two of your peers recommend the book A, the probability that you appreciate either or both recommendations is $p+(1-p)p = 2(p-p^2/2)$. If a third peer recommends book B, the probability that you will appreciate his recommendation is p. Hence, whereas you are more likely to appreciate book A because more users recommended, you are not twice as likely to appreciate it: the weight of the two peers who recommended book A went from p to p-$p^2/2$. In a more realistic model, the two users recommending book A may not be considered statistically independent, and thus, their weight would be even less. Certainly, the task of recommending to a particular individual a given product or service requires more than representativity.

McNee et al. [1] point to how some recommender systems fail because they do not produce useful recommendations. They stress that serendipity requires diversity in the recommendations—it is not useful to recommend ten highly similar products—and novelty. In other words, good recommendations maximize the flow of information [16] — ie., it must contain several distinct new elements to the individual. Hence, if a given individual is a known fan of Céline Dion, recommending the last two Dion albums is likely to be a suboptimal recommendation.

4. Do social networks achieve diversity?

We have already observed that even a large open social network may not achieve diversity. It may be difficult to measure precisely the properties of multimodal open social networks. For example, Grippa et al. [32] have shown that relying on email traces alone may underestimate democratic exchanges and overestimate the influence of a core group. It is likely that the same bias exists if we only capture one form of interaction between users. Ochao and Duval [45] have studied user-generated content from available online traces. They divided the systems into three categories. The most diverse (Amazon Reviews, Digg, FanFiction and SlideShare) have 10% of the users contribute 40% to 60% of the content. In the second category (Furl, LibraryThing and Revver), 10% of the users contributed between 60% to 80% of the content. In the last category (Scribd and Merlot) few users contributed most of the content. The first category receives a more diverseinput from users.

One indication that a network is diverse is that different people interact. According to Kelly et al. [6], discussion in political newsgroups is overwhelmingly across clusters of the like-minded, not within them. According to Fu et al. [18], blogging networks present disassortative mixing patterns: very connected bloggers tend to mix with lesser connected bloggers. Kolari et al. [28] have reported that even though internal corporate blogs are biased locally—American blogs tend to link to other American blogs—there are significant conversations across blogs from different countries as long as they share a common language (English). Kossinets and Watts [30] have suggested that, in university-based social network, homophily with respect to individual attributes such as status, gender, and age mostly has an indirect effect on the topology of the network, operating on constraints such as selection of courses and extra-curricular activities.

Another indication that a network is diverse is that it cannot be easily manipulated: we cannot find a few superinfluential people. Braha and Bar-Yam found [29] that whereas a small number of highly connected nodes can have great importance in the connectivity of the network, these hubs change quickly over time. Thus, targeting hubs may have little effect on the network. In effect, whereas a given individual may have a large impact on the network at a given time, we cannot rely on the effect to be easily reproducible. Thus diversity may arise because fast changes are possible.

Santos et al. [34] have shown that the more individuals interact, the more they must be able to adjust their partnerships quickly, otherwise cooperation becomes unprofitable and selfish interest prevails. They made their demonstration using computer simulations of the Snowdrift and Stag-hunt games, and Prisoner dilemma. If individuals can only adjust slowly their social ties, defectors (people motivated exclusively by their self-interest) eventually wipe out cooperators. If the individuals can adjust quickly their social ties, cooperators dominate. Each cooperator tends to become a hub: attracting other individuals who will benefit from the relationship. The diversity of the network as measured by the maximum number of relationships an individual can collect is maximized when rate with which individuals can change their social ties is sufficient to sustain a few cooperators, who will become popular individuals. We can imagine that at any given time, individuals may opt to become cooperators or defectors depending on their needs, objectives, and on the state of the network. While a closed social network may prevent these changes, they appear unavoidable in an open network.

The frequent occurrences of few important hubs is defining ingredient of the so-called scale-free networks. A network is said to be scale-free if the degree distribution of the nodes follow a power law. That is, the number of nodes having degree k is proportional to k^{-a} for some positive constant a. It is believed that a wide range of networks are scale-free [7,35]. Scale-free networks are also believed to be more resilient because few nodes operate at any given time as a hub whose destruction may have a noticeable impact on the connectivity of the network. Scale-free networks commonly exhibit the small-world phenomenon: whereas all nodes are not connected, we can go from any node to any other in few steps [36]. This ideas is sometimes popularized under the label "Six degrees of separation."

However, hubs and scale-free networks are not required to observe the small-world phenomenon. We say that a graph is regular if all nodes have the same degree. Using a hyper-rectangle, we are able to construct a regular graph having n nodes and log n diameter with maximal degree log n. Simply start with a 4-node rectangular graph: you have 4 nodes and a diameter of 2. Move to a 8-node cubic graph: you have 8 nodes and a diameter of 3. Generalizing this construction, you have 2d nodes and a diameter of d. In turn, a logarithmic growth of the diameter cannot be avoid: suppose that you have a graph with diameter D and maximal degree z. Starting from any node, we can reach any other node in D step, after one step we have visited at most z nodes, after two steps at most z2 nodes and so on. Hence, the graph can have at most $n = zD$ nodes or D = logz n. However, larger networks may have larger maximal degrees [37].

We define the density as the average number of links or edges per node. Intuitively, if individuals are linked to more people, they may have a more diverse input. To quantify this hypothesis, we generated 1000-node networks using Barabási-Albert preferential attachment models (Barabási and Albert, 1999). Networks are built as follows: initially, we have two nodes linked together, then we add a new node one at a time. Each node will form K edges with the existing nodes. The existing nodes are picked at random with a probability proportional to its current number of edges to some power p. In the event where a node is picked several times, only one edge is formed. For $p = 0$, we have a random graph, whereas for $p = 2$, we have a graph with strong preferential attachment: most nodes will have only K links, while very few have many more.

Hubs contribute to reduce the diameter of a network. By bringing more diverse people closer together, they may contribute to diversity. To quantify the impact of hubs on the diameter for each each value of p in 0, 1, 2 and each value of K in 1, 2, 3, 4, 5, we generated 20 random graph and computed the average diameter (see Table 1). We see that whereas strong hubs ($p = 2$) substantially reduce the diameter for a fixed density, their effect diminish quickly as the graph density increases. For a dense network ($K = 5$), there is no difference in the average diameter between having no preferential attachment and linear preferential attachment, for this particular experiment.

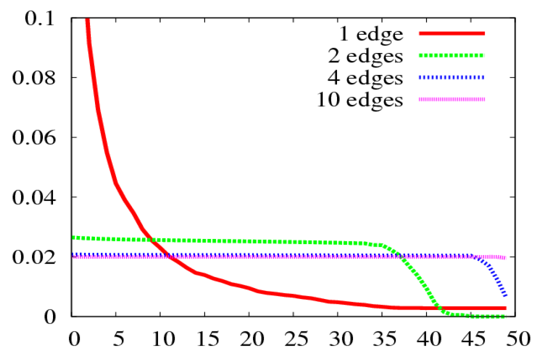
Table 1

Average network diameter with respect to the number of edges per node (K) and type of preferential attachment (p)

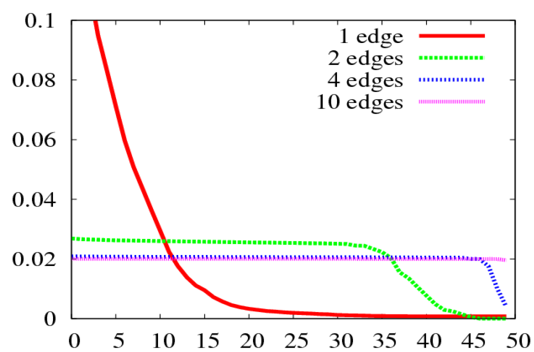
	$p = 0$	$p = 1$	$p = 2$
$K = 1$	24.8	16.8	5.1
$K = 2$	8.8	7.4	4.9
$K = 3$	6.9	6.0	4.0
$K = 4$	6.0	5.0	4.0
$K = 5$	5.0	5.0	3.9

However, the diameter might be a misleading measure of diversity: a group structure with a single leader (hub) has a small diameter, but it may not be very diverse. To better quantify diversity, we injected 50 distinct *diseases* in these networks. A disease is first seeded into one individual. Individuals pass on the disease to each of their neighbors with a probability of 50% each as long as the neighbor is not already infected. Individuals can catch several diseases, but they can catch a given disease only once. Because all diseases are initially equal, we wish them to see them infect a comparable number of people. We define the popularity of a disease as the number of people infected. This disease model can be applied to model crudely how new products or new ideas spread: if an idea are not intrinsically better, it should not unduly gain overwhelming popular as it would reduce the diversity of ideas.

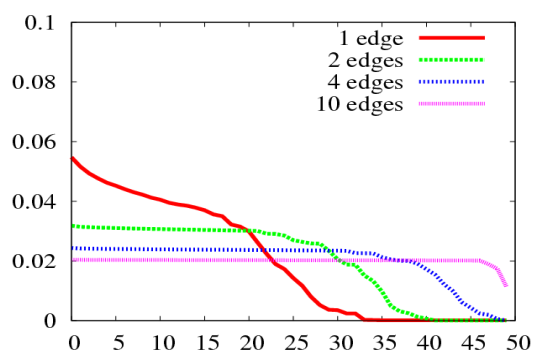
In Fig. 1, we present the relative popularity of the various diseases for different types of networks. Our results are the average of 50 different trials. We see that increasing the density of the network invariably flattens the popularity curve, a desirable property. Increasing the number of edges per node has a less significant effect on networks without strong preferential attachment (see Fig. 1.c) and is most important for networks without preferential attachment (see Fig. 1.b). In these particular experiments, the best diversity is achieved without preferential attachment, within a dense network.



a) $p = 0$ (without preferential attachment)



b) $p = 1$



c) $p = 2$

Fig. 1. Relative disease popularity relative to the number of edges per node (K) for random networks with preferential attachment parameter p

In social networks where there is almost no broadcast limit, that is, any given node can be linked to many other links, increasing the density of the graph is an efficient approach to achieve diversity. This may partially explain why Leskovec et al. [37] found that the average distance between nodes diminishes and the density increases as more nodes join networks.

4. Should social-network aggregates seek diversity?

Schwartz observed that diversity can be a source of stress [22]. Similarly, in Information Retrieval, there is a constant tension between precision—the accuracy of the result set—and recall—the completeness of the result set (akin to diversity). Fewer, but hopefully more accurate, information sources is sometimes thought to be better. For this reason, engineers commonly apply the Probability Ranking Principle (PRP): the document most likely to be relevant are presented first to the user. The PRP may actually contribute to diminish the diversity of the result set: to the query term nail, a system may produce a set of documents having to do with fingers and anatomy, but no documents about hammers and construction. In an informal test, none of the first 20 documents returned by Google using the term nail had to do with construction. However, Chen and Karger [46] have presented a new objective function for Information Retrieval which automatically enhances diversity. They seek to maximize the probability that one of the first few documents returned is relevant, instead of ranking the document by probability of relevance. Such a system would attempt to present a diverse set of documents to the users so that at least one document is relevant.

Seth [4] points out that good information must be relevant, diverse and reliable. According to Surowiecki [44], diversity of opinions leads more reliable decisions. Meanwhile, Rheingold [2] suggests that a more diverse online community increases one's chances of finding relevant information. Florida has shown that diversity is tied with innovation and prosperity [43]. Hong and Page have produced mathematical models showing that sets of diverse individuals can solve problems better than homogeneous sets of highly skilled individuals [48].

In machine learning, a common meta-learning algorithm is to combine multiple learning algorithms [17]: it works best when combining diverse learning algorithms. Mobasher et al. [19] have shown that it is possible to mount attacks against a recommender system without much knowledge of the system itself. They have shown that it is difficult to identify attacks as even a low volume of data can still have a significant impact. They argue that algorithms mixing different predictions computed in radically different ways, are more robust.

Diversity has been identified as a desirable property of recommender systems. McNee et al. [1] observe that the current breed of collaborative filtering algorithms, which often focuses on similarity between users, lead to personalized recommendations which are too predictable and ultimately not very useful to the users. Fleder and Hosanagar [20] have shown that a recommender system can actually decrease diversity. McGinty and Smyth [21] consider the role of recommendation diversity in conversational recommender systems: as you propose various options to the users, you need to ensure that these options cover a broad spectrum of choice so that the value of the user feedback is maximized. To ensure diversity, they weight the possible recommendations not only by their similarity to the user profile, but also by the dissimilarity to other recommendations.

5. Models and strategies for diversity

The long tail [8] is an hypothesis according to which e-commerce is increasing transactions of relatively unpopular items which, grouped together, form a sizeable fraction of the sales. Oestreicher-Singer and Sundararajan [31] have surveyed a large online retailer and concluded that the influence of hyperlinks is to redistribute demand between products in a way that flattens the overall distribution. In a case study, Elberse and Oberholzer-Gee [33] found that online retailing appears to shift sales towards the tail of the distribution, even though fewer star products appear to account for an even greater proportional of total sales: an even longer tail accompanies a shorter head. It does appear that the hyperlinked nature of the Web, with the multiple paths to and from any given site, helps supports the long tail.

Randomness is a practical strategy to ensure diversity while preventing gaming. As an application scenario, consider a democratic vote whereas any candidate getting more than 10,000 votes is elected. Clearly, such a system is sensitive to small cliques of similar people who choose to focus their votes on one particular candidate. We could, instead, pick the candidates at random, whereas the probability that any given candidate is picked is proportional to the corresponding number of votes. If the tail of the distribution of candidates account for 50% of the votes, then 50% of the elected candidates will be from the tail. Moreover, it is difficult for a small group to game such a system: they need to get a sizable fraction of the votes before they can have an effect. This random approach can be coupled with a weak for of thresholding: any candidate with fewer than K votes is dismissed before the random selection, and all candidates deduct K from the total of their votes. An interesting example of a robust recommender system based on randomness is the Slashdot [5] moderation system. Each registered users collect karma points over time by being active in the system in a positive way. Once a user has a sufficient amount of karma he may be randomly selected to review the comment section of a story. The user cannot transfer this moderation privilege: he must use it at once or lose the chance to moderate. It is difficult to game such a system.

Another road to diversity is representative democracy. In a social network, similar users are clustered either by their voting patterns and social links. Sets of similar users get a single vote. An equivalent view to this model is to weight original users more than conforming people: a user should be weighted according to the conditional information [42] of its votes. Simply clustering users geographically, as in traditional representative democracy may not suffice. Unfortunately, it may be difficult in practice to capture enough information in a single software applications to fully characterize clusters of similar users. Even if we were given all the information, choosing the most appropriate similarity measure may also prove to be a challenge. Similarity between users in a network may be defined topologically: two nodes are equivalent if we can exchange them without changing the structure of the network. Thus, users in a disconnected clique are all identical.

Timeliness is also a key ingredient. By making it more difficult for strong cliques and hubs to remain over time, we ensure a constant flow of new contributions. Self-reinforcing biases are diminishing diversity. However, it may not be wise to dismiss the past too quickly and become amnesic.

Ultimately, users should avoid relying on a single aggregation strategy to filter content. Indeed, diverse information sources form a more resilient system (see Section 4). For online retailers, the corresponding strategy is to offer a diverse set of aggregations to the users, including personalized ones. Retailers should seek to maximize the probability that at least some of what they offer to the users [46] is relevant, instead of simply ranking by probable relevance. Web sites like Digg offer several aggregations (technology, science, gaming and so on) to provide some measure of diversity. They should also help users grow their own personal social networks by making it easy and fast for people to interact directly, visualize, terminate and grow relations [3].

6. Conclusion and future work

It is difficult to find limitations on how useful and resilient software and recommendations based on social networks can be. However, it is easy to determine the limitations resulting from a closure of the social networks. By limiting the diversity of sources and opinions, reliability and relevance may both be diminished.

References

[1] S. M. McNee, J. Riedl, J. A. Konstan, Being accurate is not enough: how accuracy metrics have hurt recommender systems, in: Proc. CHI '06 (2006) 1097 – 1101.

[2] H. Rheingold, The Virtual Community: Homesteading on the Electronic Frontier, Addison-Wesley, Reading, 1993.

[3] A. Webster, J. Vassileva, The keepup recommender system, in: Proc. RecSys 2007 (2007) 173 – 176.

[4] A. Seth, An Infrastructure for Participatory Media, in: Proc. AAAI Workshop on Recommender Systems, 2007.

[5] C. A. Lampe, E. Johnston, P. Resnick, Follow the reader: filtering comments on slashdot, in: Proc. SIGCHI'2007 (2007).

[6] J. Kelly, D. Fisher, M. Smith, Debate, Division, and Diversity: Political Discourse Networks in USENET Newsgroups, in: Proc. Online Deliberation Conference (2005).

[7] V. V. Kryssanov, F. J. Rinaldo, E. L. Kuleshov, H. Ogawa, Modeling the dynamics of social networks, in: Proc. SIGMAP 2007 (2007).

[8] C. Anderson, The long tail: Why the Future of Business is Selling Less of More, Hyperion, New York, 2006.

[9] S. Karp, Digg Demonstrates The Failure Of Completely Open Collaborative Networks, 2008. Published online: http://publishing2.com/2008/01/24/digg-demonstrates-the-failure-of-completely-open-collaborative-networks/.

[10] M. Marshall, Knowledge raises $5M from Kleiner, on a roll, VentureBeat, 2006. Published online: http://venturebeat.com/2006/12/10/-knowledge-raises-5m-from-kleiner-on-a-roll/

[11] D. Engber, Quality Control, The case against peer review, Slate, 2005. Published online: http://www.slate.com/id/2116244/.

[12] T. Valente, Network Models of the Diffusion of Innovations, Hampton Press, Cresskill, 1995.

[13] E. Rogers, Diffusion of innovations, Free Press, Glencoe, 1995.

[14] D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in: Proc. KDD'03 (2003).

[15] P. Bourdieu, Public opinion does not exist, Communication and Class Struggle 1 (1979) 124 – 130.

[16] F. I. Dretske, Knowledge and the flow of information, MIT Press, Boston, 1981.

[17] R. Schapire, Strength of weak learnability, Journal of Machine Learning 5 (1990) 197 – 227.

[18] F. Fu, L. Liu, K. Yang, L. Wang, The structure of self-organized blogosphere, published online: http://arxiv.org/abs/math/0607361.

[19] B. Mobasher, R. Burke, R. Bhaumik, C. Williams, Toward trustworthy recommender systems: An analysis of attack models and algorithm robustness, ACM Trans. Inter. Tech. 7 (4) (2007) 23.

[20] D. Fleder, K. Hosanagar, Blockbuster culture's next rise or fall: The effect of recommender systems on sales diversity, in: Proc. WISE 2006 (2006).

[21] L. McGinty, B. Smyth, On the Role of Diversity in Conversational Recommender Systems, in: Proc. ICCBR 2003 (2003).

[22] B. Schwartz, The Paradox of Choice: Why More Is Less, Harper Perennial, London, 2005

[23] A. Gulli, A. Signorini, The indexable web is more than 11.5 billion pages, in: Proc. WWW 2005 (2005) 902 – 903.

[24] D. Benoit, D. Slauenwhite, N. Schofield, A. Trudel, World's first Class C Web census: The first step in a complete census of the Web, Journal of Networks 2 (2) (2007)

49 – 56.

[25] A. Tversky, Features of Similarity, Psychological Review 84 (1977) 327– 352.

[26] B. B. Mandelbrot, The Fractal Geometry of Nature, W. H. Freeman and Company, New York, 1982.

[27] C. E. Shannon, A Mathematical Theory of Communication, Bell Syst. Techn. J. 27 (1948) 379 – 423.

[28] P. Kolari, T. Finin, K. Lyons, Y. Yesha, Y. Yesha, S. Perelgut, J. Hawkins, On the Structure, Properties and Utility of Internal Corporate Blogs, in: Proc. ICWSM'2007 (2007).

[29] D. Braha, Y. Bar-Yam, From Centrality to Temporary Fame: Dynamic Centrality in Complex Networks, Complexity 12 (2) (2006) 59 – 63.

[30] G. Kossinets, D. J. Watts, Empirical Analysis of an Evolving Social Network, Science 311 (2006) 854 – 856.

[31] G. Oestreicher-Singer, A. Sundararajan, Network Structure and the Long Tail of Electronic Commerce, Working Paper, New York University, August 2006.

[32] F. Grippa, A. Zilli, R. Laubacher, P. Gloor, E-mail May Not Reflect The Social Network, in: Proc. International Sunbelt Social Network Conference (2006).

[33] A. Elberse, F. Oberholzer-Gee, Superstars and Underdogs: An Examination of the Long Tail Phenomenon in Video Sales, Harvard Business School Working Paper Series, No. 07-015, 2006.

[34] F. C. Santos, J. M. Pacheco, T. Lenaerts, Cooperation prevails when individuals adjust their social ties, PLOS Computational Biology 2 (2006) 1284 – 1291.

[35] A.-L. Barabási, Scale-Free Networks, Scientific American 288 (2003) 60 – 69.

[36] D. J. Watts, S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440 – 442.

[37] J. Leskovec, J. Kleinberg, C. Faloutsos, Graph evolution: Densification and shrinking diameters, ACM Transactions on Knowledge Discovery from Data 1 (1) (2007) 2.

[38] A. Forte, A. Bruckman, Why do people write for Wikipedia? Incentives to contribute to open-content publishing. in: Proc. GROUP workshop (2005).

[39] C. Wagner, P. Prasarnphanich, Innovating Collaborative Content Creation: The Role of Altruism and Wiki Technology, in: Proc. Annual Hawaii International Conference on System Sciences (2007) 18.

[40] C. G. Wu, J. H. Gerlach, C. E. Young, An empirical analysis of open source software developers' motivations and continuance intentions, Information & Management 44 (3) (2007) 253 – 262.

[41] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509 – 512.

[42] S. E. Hodge, The information contained in multiple sibling pairs, Genet. Epidemiol. 1 (2) (1984) 109 – 122.

[43] R. L. Florida, The rise of the creative class, Topeka Bindery, 2004.

[44] J. Surowiecki, The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations, Anchor, Harpswell, 2004.

[45] X. Ochoa, E. Duval, Quantitative analysis of user-generated content on the Web, in: Proc. Web Science workshop WebEvolve (WWW 2008) (2008).

[46] C. Harr, D. R. Karger, Less is more: probabilistic models for retrieving fewer relevant documents, in: Proc. SIGIR'06 (2006) 429 – 436.

[47] Levy, Paul S. and Stanley Lemeshow, Sampling of Populations: Methods and Applications, Third Edition, John Wiley and Sons, Inc., New York, 1999.

[48] L. Hong, S. E. Page, Groups of diverse problem solvers can outperform groups of high-ability problem solvers, Proceedings of the National Academy of Science 101 (46) (2003) 16385 – 16389.