The role of (non-)conformism in rating platforms

Chih-Chun Chen* and Camille Roth *,†
*CAMS (Centre d'Analyse et de Mathématique Sociales), CNRS/EHESS
190 Avenue de France, 75013 Paris, France
†ISC-PIF (Institut des Systèmes Complexes - Paris Île de France)
57-59 rue Lhomond, 75005 Paris, France
E-mails: c.chen@abmcet.net, roth@ehess.fr

***Abstract*—Social content-sharing networks allow users to share content annotations. Although convergence and consistency in semantic annotation (tags) has been well-studied, less effort has been devoted to studying evaluative annotations (ratings and reviews) with respect to user characteristics and user-item relationships. In this paper, we first identify trends in both item scores and in the ways in which users allocate scores; these are also associated with some of users' other activity – in particular, social links and rating activity are both found to be higher for a moderate level of non-conformism and dissensus. We then conduct a more thorough investigation into how item score distributions might arise and be sustained. The fact that most items have a clear modal score can be attributed to the tendency of items to evoke similar degrees of satisfaction across users. In addition to this however, our findings suggest that social links between users can play a role in stabilising rating distributions. Firstly, we find that socially linked users are more likely to give the same score to an item (possibly due to similarities in taste). Secondly, we eliminate the possibility that distributions of scores arise through attracting users with particular ratings styles (e.g. tendency to agree). Thirdly, we find that a large mean shift is much rarer for items with a large proportion of added scores from socially linked users and that this is most likely to be due to maintaining a stable distribution than to added scores conforming to the mean.**

I. Introduction

An important function served by content-sharing social networks is to allow users to share content and content annotations. Indeed, a successful content-sharing social network can be defined as one where members continuously contribute content and collectively annotate it in such a way that for every member, the content made available or recommended to him/her is content that satisfies his/her preferences.

Although a significant body of work has addressed semantic annotation (tags), less attention has been paid to evaluative annotations (ratings) in the context of a social network platform, especially with respect to user characteristics. While community convergence in semantic annotation allows us to place an item in some semantic space, for many rating systems it is not immediately clear that evaluative annotations do the same for the item in 'taste' space (unless of course the rating system explicitly defines the semantic dimensions being evaluated, e.g. [1]). In many online communities, the rating system is only a one-dimensional, undefined scale. For a given item, two users might give scores at opposite ends of the rating scale in response to the very same intrinsic property (e.g. 'overly sentimental', score 1 vs. 'touching', score 4). And different users can give the same score for very different reasons. In addition, users can differ in their rating behaviour (for example, some users may try to conform to the ratings of their friends or the ratings they see; others may only rate when they disagree with the existing mean rating).

A. Related work

Collaborative filtering recommender systems are often based (at least partly) on user rating systems [2], [3] which are used to build similarity profiles between users. Users who share many items and rate them similarly are believed to have similar tastes, so that a high rating of an item by one of the users can be used as a basis for recommending this item to the second user. A major issue in collaborative filtering is the sparsity of shared item ratings, which means that it usually has to be enhanced with content-based information [4], [5]. More recently, measures taking into account the non-independence of relationships between users and objects have been introduced, such as the generalised model of relational similarity [6], which is based on the fact that similar objects are also rated similarly by similar actors.

An issue that has been troubling those studying online communities and social networks is distinguishing the effects of homophily, social influence, and external common causes [7], [8], [9]. In the current paper, we do not seek to identify causal relationships between social links and ratings but to determine whether or not the data are consistent with there being underlying taste commonalities between users.

Apart from these issues of user and item similarities, several relatively recent studies focused on the nature of reviews themselves. As regards quantitative evaluations, [10] study ratings on a dataset provided by *amazon.com* where they conclude that item ratings are essentially bimodal; they attribute this to an underrepresentation of moderate reviews. Another stream aims at predicting future ratings from the structure of existing ones, such as using various aggregation and/or machine-learning methods as in [11], or carrying regressions on the content of reviews, using for instance review length and rating evaluations as in [12], sometimes using full text models which go beyond "bags of words" or document vectors [13]. Finally, some authors addressed the credibility of reviews by examining which factors and conditions reinforced review impact [14], [15] – i.e., in short, the issue of meta-rating: how ratings are subjectively rated. In general, these works examine rating distributions independently of (platform-wide) user characteristics or user-item relationships.

B. Outline

The section that follows gives a brief description of the dataset and our methodology. Section III then analyses the differences between users' rating styles in terms of both tendency to diverge and in terms of the polarity of their scores. In Section IV, we show that social connections are associated with agreement and that this is a plausible mechanism for maintaining the stability of rating distributions.

II. Dataset and methodology

The data analysed in this study comes from the book rating platform aNobii. It was collected by the authors of [16], who performed an analysis of the interaction between social network evolution and profile similarity. The authors took six snapshots of the network 15 days apart. In our study we used only the first and fourth snapshots (hence 45 days apart) since per book, change in ratings was slow; we excluded the fifth and sixth snapshots where platform administrators had changed the rating scale from a "1-4" range to a "1-5" range.

In aNobii, the number of ratings per item and number of items rated per user exhibit heterogeneous, power-law like distributions (Figure 1: left). The distribution of items rated by users has an initial slow decay followed by a sharper one (Figure 1: right).

So as to avoid statistical artifacts arising from over-representation of distributional differences for items with small numbers of ratings (for example, agreement is over-represented), we consider only books with 10 or more ratings (hereon we shall refer to these as 'popular books'). This represents around 12% of the books (98 995 out of 847 984), and around 74% of the ratings (4 574 764 ratings out of 6 115 183). 60 093 distinct users rated books in this set, and of these, 31 362 rated 10 or more books in the set. So as to have confidence in distributional analyses at the user level, we consider only this latter set of users and will hereon refer to them as the 'rating community'.

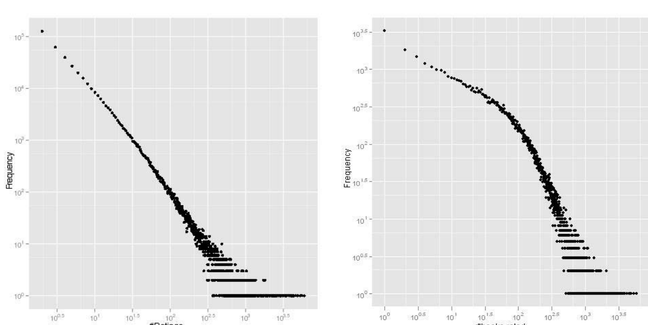


Fig. 1. Left: Frequency plot of number of ratings per book. Right: Frequency plot of number of items rated per user.

General trends

In order to put our analyses in context, it is worth first highlighting some key features of the rating community:

- Regarding ratings:
  - 3 and 4 scores have a tendency to dominate for raters (Figure 2(a)),
  - agreement is relatively common, and rating distributions tend to be unimodal, with little divergence from the single mode, as exemplified by Figure 3 which represents the aggregate distribution, over all books, of rating divergence with respect to the median rating of each book.

    These results diverge from [10] who show that raw ratings on Amazon.com are bimodal, and who suggest that this bimodality is due to bipolar/extreme ratings. In other words, the aNobii platform seems to prevent this rating style.[1]
- Regarding sociability, our dataset displays characteristics similar to those reported earlier by [16]:
  - Social links are rare, with an overall density of $1.9 \cdot 10^{-4}$ for the union of friendship and neighbour networks; this is higher than reported for the whole user-base in [16] ($9.3 \cdot 10^{-5}$), most likely because the subset of users in our dataset are also likely to be more active and hence more social. Connectivity is also weak, with an average out-degree of 10.75 (though slightly higher than for network of the whole user-base, which has an average out-degree of ≈ 8 [16]).
  - Reciprocal links are more common than one-way links (58% for the union of friendship and neighbour networks for raters of popular book raters, only slightly higher than that for the whole network, 57%).

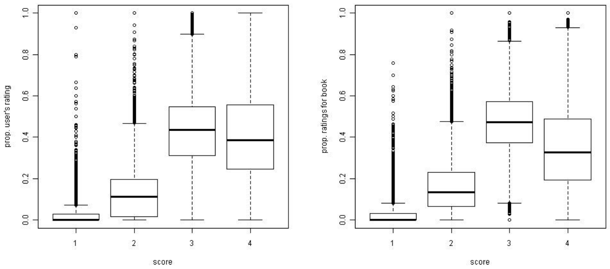
Since agreement between pairs of users is common (and deviations tend to be small at the item level) and links between users are rare, analyses based on linear relationships are unlikely to yield much insight. For this reason, the majority of our analyses try to identify differences between classes or communities of users, rating behaviour and items.

Items exhibit a similar pattern to that observed for users in terms of rating distribution, with the majority of items having highest proportions of 3 and 4 scores (Figure 2(b)).

III. Individual differences in rating styles

It is important to distinguish between ratings and rating *style*. The score that a user comes to give to an item can be seen as the combined outcome of both his/her response to the item (taste), and his/her typical rating style, which determines how this response is manifest. For example, a user with a tendency to give extreme scores may express his disappointment with an item with a score of 1 where another user with more moderate score assignment would assign a

[1] We may perhaps speculate that this discipline is caused by some community aspect of the platform.



(a) By user (b) By book

Fig. 2. Distributions of scores.

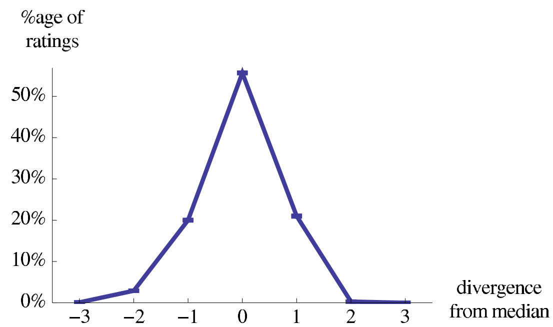


Fig. 3. Distribution of ratings translated with respect to the median rating of each book, and averaged all books (Note: for the sake of simplicity, these distributions are computed on books whose ratings median is an integer, which represents around 95.5% of books with more than 10 ratings.)

score of 2. Although we found variety among users in their distributions of scores, this can be difficult to interpret. For example, users giving high proportions of their items a score of 1 might be severe in the way they rate, but they might also have been more unlucky in their choice of books.

The rating that a user eventually gives an item can be seen as a function of both taste and the user's typical rating behaviour (rating style) (e.g. tendency to give extreme scores). As with tagging [17] and contribution in general to online communities [18], users may well differ in their motivations for rating and/or the way they use the ratings system [19]. Some users may for instance only rate an item when it has evoked a strong response, some might rate to keep a personal record of how much they have enjoyed an item and so rate nearly every item they have read, and still others may only add their rating if it differs significantly to those already existing. Therefore when talking of rating behaviour, we do not give any speculative interpretation of what is giving rise to these individual tendencies. Rather, we focus on identifying associations between these tendencies and users' other activity.

A. *Users with robust rating behaviours*

As noted in Section II, for a given item, there tends to be consensus, with ratings clustering around a modal score, which can be seen as a community-generated evaluation of the item. Within this context however, a given user can show a high degree of divergence.

To get an idea of how members of the rating community vary in terms of their tendency to diverge, we computed the means of absolute and signed divergence of scores across all a user's n items expressed in terms of item standard deviations for all members of the rating community. More formally, for a given user with n ratings we define:

$$\overline{d_{\text{abs}}} = \frac{\sum_{i=0}^{n} |d_i|}{n} \quad \text{and} \quad \overline{d_{\text{sig}}} = \frac{\sum_{i=0}^{n} d_i}{n} \tag{1}$$

where

$$d_i = \frac{x_i - \mu_i}{\sigma_i} \tag{2}$$

($\overline{d_{\text{abs}}}$ and $\overline{d_{\text{sig}}}$ correlated strongly with their median equivalents; r=0.913 and hence did not appear to be sensitive to bias by extreme values.)

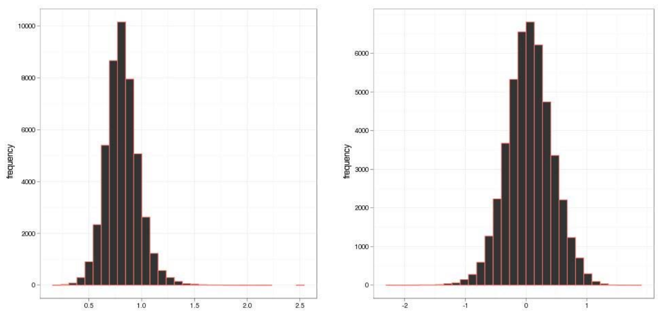
Absolute divergences indicate how much a user consistently holds an opinion different from the mean (as such a marker of 'non-conformism'), while signed divergences indicate whether this opinion consistently tilted towards a certain direction (as such a marker of negative or positive 'bias'). Figure 4 shows the distribution of $\overline{d_{\text{abs}}}$ and $\overline{d_{\text{sig}}}$ across users. Both the distribution of absolute and signed factors appear to be clustered around a central value (≈ 0.8 and ≈ 0.05 respectively). Consistent with the tendency to give more positive scores, the distribution of signed divergence factors is slightly negatively skewed (skewness -0.05); since raters have a tendency to rate positively, on average those who diverge more will also be those who diverge negatively.

In what way may the rating style be related to conformism or bias? In other words, do non-conformist (high $\overline{d_{\text{abs}}}$) or severe (negative $\overline{d_{\text{sig}}}$) users use a particular palette of ratings, compared with other users? To check this, we examine the association between the mean and spread of their ratings: Figure 5 plots the means and standard deviations of users' raw scores, against the means of their absolute and signed item divergences.

There is a positive association between the mean of users' raw ratings μ_{raw} and $\overline{d_{\text{sig}}}$, and between the standard deviation of their raw ratings σ_{raw} and $\overline{d_{\text{abs}}}$:

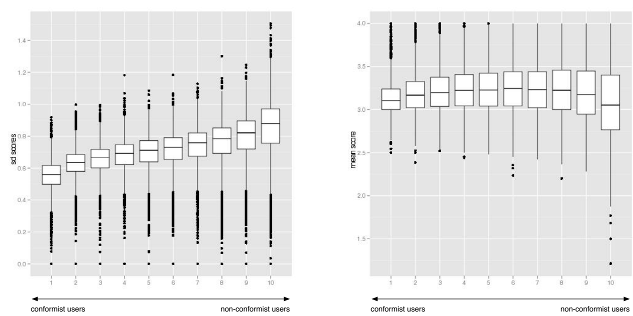
- The association between $\overline{d_{\text{abs}}}$ and σ_{raw} (Figure 4(a): left) suggests that 'non-conformism' is associated with having more different scores, although the last decile also seems to have greater variation in σ_{raw}. The last decile also has lower μ_{raw} (Figure 4(a): right), implying a tendency for these non-conformists to diverge negatively.
- The association between μ_{raw} and $\overline{d_{\text{sig}}}$ (Figure 4(b): right) suggests that users are in general equally likely to rate 'high quality' items (items with high means) as they are poorly rated ones (therefore users who have a tendency to rate more positively also have higher means, and those with a tendency to rate more negatively tend to have lower means; a lack of association would have suggested that users diverging positively were rating more 'poor quality' items and/or those diverging negatively were rating more 'high quality' items).

The final three deciles of $\overline{d_{\text{sig}}}$ appear to exhibit a negative trend with respect to σ_{raw} (Figure 4(b): left), suggesting that those who tend to diverge more positively show less variation in their scores. This seems to suggest that consistency is stronger on the positive end of the spectrum.

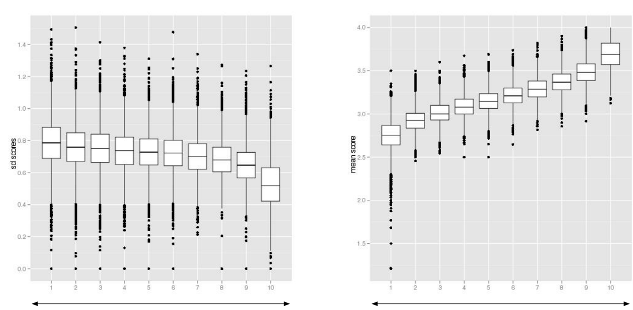


(a) $\overline{d_{\text{abs}}}$ (absolute values) (b) $\overline{d_{\text{sig}}}$ (signed values)

Fig. 4. Distribution of user divergences.



(a) $\overline{d_{\text{abs}}}$

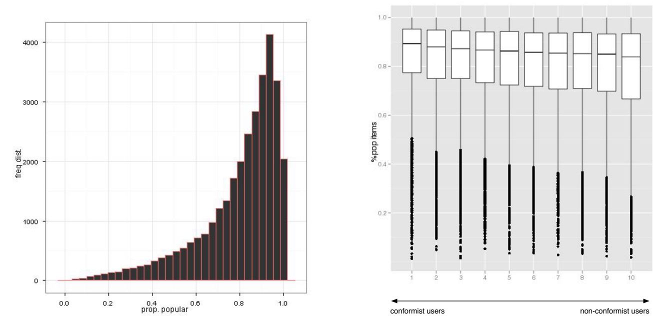


(b) $\overline{d_{\text{sig}}}$

Fig. 5. Boxplots of $\overline{d_{\text{abs}}}$ and $\overline{d_{\text{sig}}}$ deciles against the standard deviations (left) and means (right) of raw user scores.

B. Rating behaviour and book popularity

Of the 31362 users rating 10 or more popular books, 1777 rated only popular books, while the overwhelming majority of 29585 also rated less popular books. Not surprisingly though, for most users, a larger proportion of their ratings were for popular books (see Figure 6). More interestingly, there is a small but significant association between the proportion of popular books a user rates and his/her behaviour in terms of



(a) Distribution of proportion of popular books among the books users rate

(b) Boxplot of proportions of popular book split into $\overline{d_{\text{abs}}}$ deciles (i.e. vs. 'non-conformism')

Fig. 6. Rating behaviour and book popularity

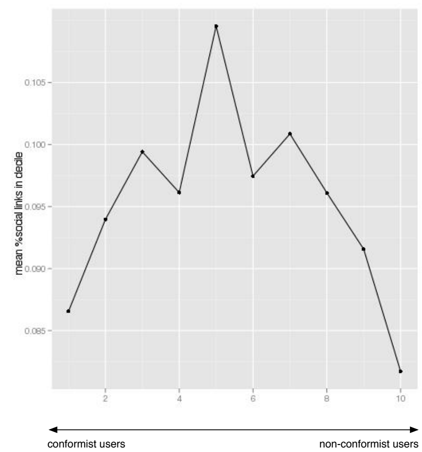


Fig. 7. Mean proportion of users' social links that come from within the same $\overline{d_{\text{abs}}}$ decile.

divergence, in the sense that more conformist users tend to have a slightly higher proportion of popular books, on average.

C. Rating behaviour and social activity

In terms of social activity, raters at both extremes of the divergence scale (top and bottom deciles of $\overline{d_{\text{abs}}}$) tended to be less social, with both fewer friends and fewer neighbours. They also had a slightly lower likelihood of being socially connected to each other (even taking into account the lower number of connections overall); this implies that a lower proportion of those socially linked to them had the same rating habits. The 5th decile was both the most social, with a mean of 26.2 links per user (compared to 19.6 for the first decile and 19.3 for the last decile) and had a slightly higher proportion of links coming from other users in that decile. (See Figure 7.)

IV. Score distributions as the outcome of community rating

A. Items as communities of raters

Just as there are individual differences among raters in the way they typically rate, there may also be differences among items in the types of raters they attract. Although part of this question is answered by the rating distributions, the answer is not complete. For example, it is possible for an item to show less clustering around its central score, but the raters of this item are not necessarily those who typically diverge in their ratings (as measured by their individual-level divergences, as described in Section III-A); its ratings could instead be dominated by raters who usually do not diverge much but do for this item. For each item, the mean of users' $\overline{d_{\text{sig}}}$ can be used as a measure of the polarity of users' responses; while the mean of users' $\overline{d_{\text{abs}}}$ as a measure of users' tendency to diverge. Formally, for an item with m ratings, we may define:

$$D_{\text{abs}} = \frac{1}{m}\sum_{j=0}^{m}\overline{d_{\text{abs}}(j)} \quad \text{and} \quad D_{\text{sig}} = \frac{1}{m}\sum_{j=0}^{m}\overline{d_{\text{sig}}(j)} \qquad (3)$$

If the association between D_{sig} and the item rating mean μ were high, this would imply that positively rated items tended to be those attracting positive users and negatively rated items tended to attract negative/unlucky users. Similarly, if the association between D_{abs} and the item rating standard deviation σ were high, this would suggest that variety in scores could be attributed to more raters having more 'extreme' rating behaviours.

The data did not support either of these. On the contrary, we found the association between d_{abs} and σ to be rather weak (though still significant at $p = 0.001$), $r^2 = 0.011$. The same was true between item μ and D_{sig}, $r^2 = 0.015$ ($p < 0.001$). This seems to suggest that in general the distribution of item ratings is not the outcome of users with particular rating styles tending to rate the same items (e.g. raters with a tendency to diverge preferring more 'controversial' books). Instead, items seem equally likely to bring out users' typical and atypical rating behaviours.

These findings contrast with those for the distribution of scores for each user, where users who tend to rate more positively across their items (some of which may have low means) also tend to have higher means, and raters who tend to diverge across their items also show more deviation amongst their own scores (i.e. they are not strongly polarised in their divergence).

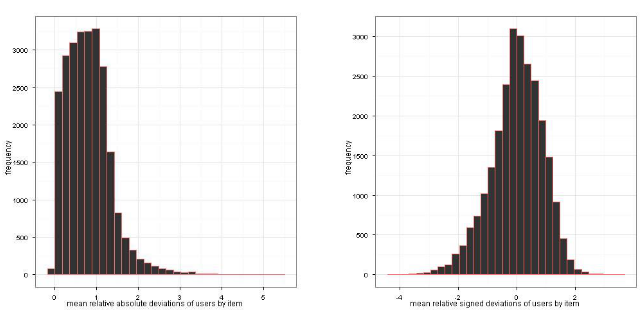
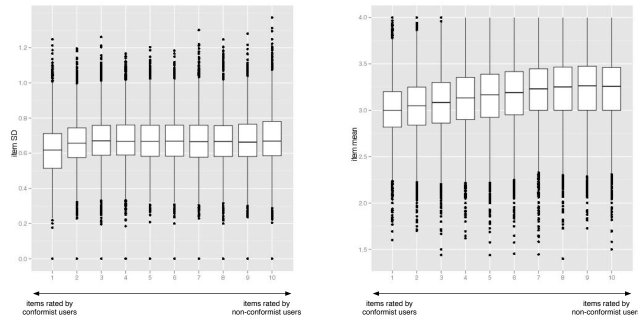
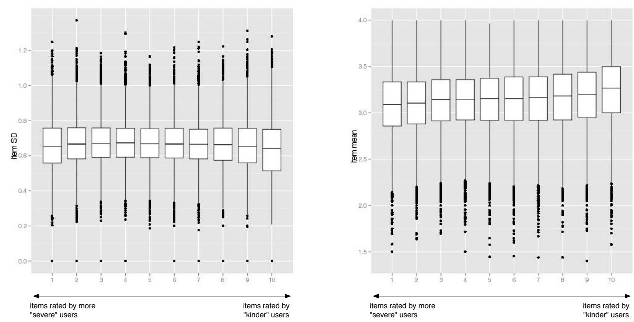


Fig. 8. Left: Item means of the relative deviations of users' ratings with respect to their personal mean deviations. Right: Item means of the relative signed deviations of users' ratings with respect to their personal mean deviations.



(a) $\overline{D_{abs}}$



(b) $\overline{D_{sig}}$

Fig. 9. Boxplots of item score standard deviations ("SD", left) and means (right) grouped by deciles of the mean absolute (top) and signed (bottom) deviations of the users rating them.

B. Agreement and social links

Although as already pointed out in Section II social links are in general relatively rare and agreement in ratings is relatively common, if the likelihood of agreement is higher between socially linked users, this could still play a significant role in shaping the rating distribution. (This need not be given a causal interpretation; indeed, social links are also likely to be associated with similarity in taste, which is the more likely reason for agreement than direct social influence.)

In terms of the proportion of pairs agreeing, the difference between socially and non-socially linked users was significant but small (solid line in Figure 10). This is in large part due to the high levels of agreement found for most items and the infrequency of social links.

To identify cases where social links (perhaps indicating underlying taste) would be more likely to make a difference, we took pairs of raters who agreed most with each other but disagreed most with the mode rating. For example, if the mode rating was 3, we took two of the raters giving a rating of 1. We also took into account the overall degree of consensus of an item, which we measure by the standard deviation σ_b for an item b. Given that distributions tend to be unimodal (see Figure 3), σ_b gives a good indication of the distribution's peakedness. The second and last deciles were then taken to represent respectively 'high consensus' and 'low consensus' items. (The second decile was used instead of the first because, although the two are qualitatively indistinguishable, a large number of books in the first had agreement for all possible pairs.) Figure 10 shows the proportions of social links for pairs with different agreement levels in high and low consensus items. For comparison, we also included the sixth decile as ‘mid consensus’ items. It is worth noting that these mid consensus items had roughly twice as many raters than low consensus or high consensus items, suggesting that a moderate level of disagreement is associated with more rating activity.

We predicted that the difference between socially linked and non-socially linked pairs would be greater for low consensus books since for these books, the likelihood of a given pair being different from other possible pairs would be lower and hence the association between social links and agreement more easily identified.

For low consensus (high values of σ_b) items, a high agreement rate of 0.638 (3sf.) was found in pairs of raters who were neighbours, which was significantly higher (at the $p < 0.001$ level) than the agreement rate for non-neighbour pairs (0.499). For consensual items, the effect was weaker (though still significant at $p < 0.01$), with an agreement rate of 0.574 between neighbours and 0.499 between non-neighbours. We also found that agreement is more common between friends than non-friends (though the effect is weaker than for neighbours) and that disagreement is less common among friends than non-friends and that this effect is greater for low consensus items than for high consensus items.

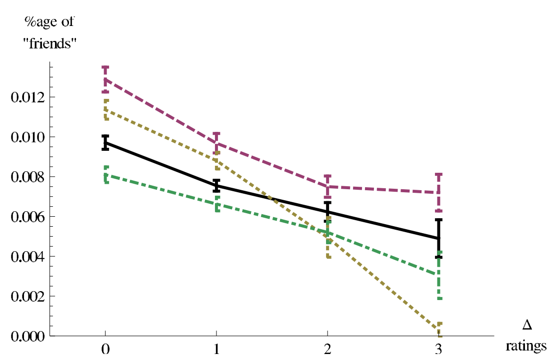


Fig. 10. Proportion of pairs with social links for each level of agreement (Δ ratings) and for the various levels of consensus for book ratings (solid line: average over all books, dashed: low consensus, dot-dashed: mid consensus, dotted: high consensus).

C. Social links as a mechanism for stabilising rating distributions

Given that social links imply a greater likelihood of agreement in scores, we hypothesised that if a large proportion of an item’s raters are socially linked to each other, the item should have a more peaked (distinctly unimodal) distribution. We found no such relationship. However, when we considered ratings over time (between the first and fourth time snapshots in the data provided by [16]), we found that when a large proportion of scores were added by users who were socially linked to those who had previously rated the item, the item’s mean was less likely to change by large values.

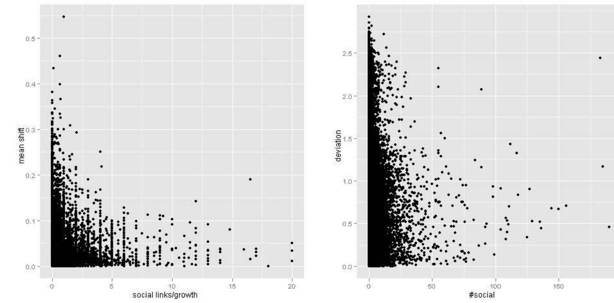


Fig. 11. Left: Scatter plot of mean shift against social links/growth. Right: Scatter plot of user deviations against the number of social links existing in the item they have rated.

As social links are rare, when considered globally, the proportion of additional ratings coming from users who were socially linked to users who had already rated an item made up only around 10% of the additional ratings. However, for some items, a large proportion of added scores came from socially linked users. Large mean shifts were less frequent for these items (Figure 11, left), implying that when a large proportion of new raters are socially linked to existing ones, the book’s mean is more likely to remain stable.

One explanation for this is that raters who are driven to read and rate a book by one or more of their socially connections is more likely to rate close to the mean. An alternative explanation is that it is the *distribution* of scores that is maintained. This would imply that the rates at which each of the score occur remains relatively constant over time due to the frequency of social links being proportional to the number of users giving that score.

At the user level, no general relationship was found between deviation from the mean rating and the number of social links between added and previous raters (Figure 11, right). In other words, a rater contributing a score in the later time slice who was socially linked to a user who had previously rated the item was not any more likely to rate close to the mean than one who had no social link with a user previously rating the item.

Consistent with this, at the item level, for low numbers of social links, there appeared to be little association between the change in mean and the existence of social links; items for which a higher proportion of the additional raters socially linked to existing raters were not any more likely to retain the same distribution (since most books tended to retain similar distributions around the same central value).

Our findings therefore suggest that it is the distribution of ratings that is sustained through the greater tendency for agreement between socially linked users.

V. Summary and Conclusions

The goal of this paper was to identify trends in the evaluation of items in an online community and to probe more deeply into the mechanisms underlying these. In particular, we wished to establish how item rating distributions and average ratings arise from communities of socially linked users. Our analyses were conducted on data from the online book-sharing site aNobii, but we envisage the protocols introduced to also be applicable to other online communities with rating systems.

For the aNobii users we studied, we found individual differences in rating styles, with some users having a greater tendency to diverge from the central score and others being more likely to conform. These were also associated with some of users' other activity. For example, both highly conformist and highly divergent users had fewer social links, implying that there is an optimal intermediate level of conformism that associates positively with activity. Similarly, items with an intermediate level of consensus tended to have more raters, suggesting that discussion tends to be facilitated when there is some, but not too much disagreement. Aside from social psychology, these features could be helpful to the design of recommender systems.

We then tried to identify the mechanisms by which item score distributions might arise and be sustained. For the items we studied, the large majority had a modal score, which can be seen to represent the community evaluation of the item. This can be attributed to the tendency of items to evoke similar degrees of satisfaction across users (even if this might be for different reasons).

In addition, the set of scores for an item can be generated by users who already have a higher likelihood of giving similar scores. Our findings suggest that social links can provide a means for keeping both the central tendency and distribution of scores stable. Firstly, we find that socially linked users are more likely to give the same score to an item (possibly due to similarities in taste). Secondly, we eliminate the possibility that distributions of scores arise through attracting users with particular ratings styles (e.g. tendency to agree). Thirdly, we find that a large mean shift is much rarer for items with a large proportion of added scores from socially linked users and that this is more likely to be due to maintaining a stable *distribution* of scores than to added scores converging to the mean.

Acknowledgments. The authors would like to thank Fred Amblard for the fruitful discussions and Luca Mario Aiello for his support on using the data set. This work was partially supported by the Future and Emerging Technologies programme FP7-COSI-ICT of the European Commission through project QLectives (grant no.: 231200).

References

[1] S. Aciar, D. Zhang, S. Simoff, and J. Debenham, "Recommender System Based on Consumer Product Reviews," in *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, ser. WI '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 719–723. [Online]. Available: http://dx.doi.org/10.1109/WI.2006.144

[2] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Commun. ACM*, vol. 35, no. 12, pp. 61–70, Dec. 1992. [Online]. Available: http://dx.doi.org/10.1145/138859.138867

[3] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Trans. on Knowl. and Data Eng.*, vol. 17, no. 6, pp. 734–749, Jun. 2005. [Online]. Available: http://dx.doi.org/10.1109/TKDE.2005.99

[4] J. Basilico and T. Hofmann, "Unifying collaborative and content-based filtering," in *Proceedings of the twenty-first international conference on Machine learning*, ser. ICML '04. New York, NY, USA: ACM, 2004, pp. 9+. [Online]. Available: http://dx.doi.org/10.1145/1015330.1015394

[5] P. Melville, R. J. Mooney, and R. Nagarajan. (2002) Content-Boosted Collaborative Filtering for Improved Recommendations. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.73.8546

[6] B. Kovács, "A generalized model of relational similarity," *Social Networks*, vol. 32, no. 3, pp. 197–211, Jul. 2010. [Online]. Available: http://dx.doi.org/10.1016/j.socnet.2010.02.001

[7] A. Anagnostopoulos, R. Kumar, and M. Mahdian, "Influence and correlation in social networks," in *In Proc. of the 14th ACM Int. Conf. on Knowledge Discovery and Data Mining (KDD'08)*, 2008. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.159.2066

[8] S. Aral, L. Muchnik, and A. Sundararajan, "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks," *Proceedings of the National Academy of Sciences*, vol. 106, no. 51, pp. 21 544–21 549, Dec. 2009. [Online]. Available: http://dx.doi.org/10.1073/pnas.0908800106

[9] C. R. Shalizi and A. C. Thomas, "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies," *Sociological Methods and Research*, vol. 40, pp. 211–239, Nov. 2011. [Online]. Available: http://arxiv.org/abs/1004.4704

[10] N. Hu, P. A. Pavlou, and J. Zhang, "Can online reviews reveal a product's true quality? empirical findings and analytical modeling of online word-of-mouth communication," in *Proc. EC'06 7th ACM Conf. Electronic Commerce*, 2006, pp. 324–330.

[11] M. McGlohon, N. Glance, and Z. Reiter, "Star quality: Aggregating reviews to rank products and merchants," in *Proc. 4th ICWSM Intl AAAI Conference on Weblogs and Social Media*, 2010.

[12] S. M. Mudambi and D. Schuff, "What makes a helpful online review? a study of customer reviews on amazon.com," *MIS Quarterly*, vol. 34, no. 1, pp. 185–200, 2010.

[13] N. Archak, A. Ghose, and P. G. Ipeirotis, "Show me the money! deriving the pricing power of product features by mining consumer reviews," in *Proc. 13th ACM SIGKDD Intl Conf. Knowledge discovery and data mining*, 2007, pp. 56–65.

[14] P. yu Chen, S. Dhanasobhon, and M. D. Smith, "All reviews are not created equal: The disaggregate impact of reviews and reviewers at amazon.com," 2008, available at SSRN: http://ssrn.com/abstract=918083.

[15] C. Danescu-Niculescu-Mizil, G. Kossinets, J. Kleinberg, and L. Lee, "How opinions are received by online communities: A case study on amazon.com helpfulness votes," in *Proc. WWW'09 Intl World Wide Web Conference*, 2009.

[16] L. M. Aiello, A. Barrat, C. Cattuto, G. Ruffo, and R. Schifanella, "Link creation and profile alignment in the aNobii social network," in *SocialCom '10: Proceedings of the Second IEEE International Conference on Social Computing*, Minneapolis, Minnesota, USA, Aug. 2010, pp. 249–256.

[17] V. Mirzaee and L. Iverson, "Tagging: Behaviour and motivations," *Proc. Am. Soc. Info. Sci. Tech.*, vol. 46, no. 1, pp. 1–5, 2009. [Online]. Available: http://dx.doi.org/10.1002/meet.2009.14504603122

[18] G. Beenen, K. Ling, X. Wang, K. Chang, D. Frankowski, P. Resnick, and R. E. Kraut, "Using social psychology to motivate contributions to online communities," in *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, ser. CSCW '04. New York, NY, USA: ACM, 2004, pp. 212–221. [Online]. Available: http://dx.doi.org/10.1145/1031607.1031642

[19] A. M. Rashid, K. Ling, R. D. Tassone, P. Resnick, R. Kraut, and J. Riedl, "Motivating participation by displaying the value of contribution," in *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*. New York, NY, USA: ACM, 2006, pp. 955–958. [Online]. Available: http://dx.doi.org/10.1145/1124772.1124915\%3C