Detecting Spammer Communities Using Network Structural Features

Wen Zhou, Meng Liu(✉), and Yajun Zhang

School of Computer Engineering and Science,
Shanghai University, Shanghai 200444, China
zhouwen@shu.edu.cn, saraliu1994@gmail.com, zyj1985email@163.com

Abstract. Spammers generate fake reviews to influence the reputation of products. By grouping together, spammers can dramatically alter how products are perceived. Different from previous research, which has mostly used behavioral indicators and structural indicators, we propose a new perspective on spammer detection. In our approach, we portray reviewers as a comment-based reviewer network through a new collusion similarity measure, divide reviewers into different communities using an effective community detection method and separate spammer communities from normal reviewer communities through network structure. We find that spammer communities have different network structural features from normal reviewer communities, a high clustering coefficient and high self-similarity. In our experiments, we show that our method achieves a detection accuracy of 94.59% - substantially higher than the current state-of-the-art methods which achieve an 80.00% accuracy.

Keywords: Fake reviews · Spammer community
The comment-based reviewer network · Network structural features

1 Introduction

With the rapid development of social networks, the fake information of the platform can bring great harm [12]. Spammers on e-commerce platforms such as Amazon, Yelp or the large Chinese website Dianping, are paid to boost or libel products by generating a large number of fake reviews. In 2015, Amazon sued 1,114 spammers who were paid $5 each to provide fake 5 stars reviews for products. Fake reviews make review data lack authenticity and create barriers for follow-up research. This problem must be solved effectively.

The first technique for spammer detection was proposed by Jindal et al. [2]. Many papers e.g. [1,3,4], identify spam or individual spammers by content analysis. Mukherjee et al. [6] were the first use behavioral indicators of fake reviews to detect spammers. Since then, behavioral indicators have become an important basis for spammer detection. Besides this work, few papers, e.g. [7], [10], identify spammer communities. The problem of spammer community detection was first proposed by Mukherjee et al. [7]. They ranked the "spamicity" of each community using several behavioral indicators (BIs). Generally, spammer community

I. Romdhani et al. (Eds.): CollaborateCom 2017, LNICST 252, pp. 670–679, 2018.
https://doi.org/10.1007/978-3-030-00916-8_61

detection is largely based on BIs. However, the method proposed by Want et al. [10] constructs the reviewer network as a bipartite graph, and discovers spammer communities by structural indicators (SIs).

Unfortunately, spammers have become extremely sophisticated. The behavior of spammers has become more hidden and it is more difficult to identify spammers only by using BIs or SIs. To make matters worse, there are more spammers working together in spammer communities. Compared with individual spammers or spam, a spammer community is more harmful because of the use of clever technology. At this point, a single index can't effectively detect spammer communities, and previous spammer detection methods encounter unprecedented challenges.

Coincidentally, review data can be represented as a comment-based reviewer network. Hence, spammer community detection can be solved by using the theories and methods developed in the research on complex networks [13]. In order to effectively distinguish spammer communities from normal reviewer communities, this paper applies a community detection method to spammer community detection. We find that previous research ignores the network structure of spammer communities. To identify spammer communities, this paper uses self-similarity (SS) and clustering coefficient (CC) as network structural features (NSFs). In terms of NSFs, the internal links of a spammer community are dense and different from a normal reviewer community, and the similarity among spammers in a spammer community is extremely high. For the case of Dianping (a popular website in China), this paper shows that CC has the best performance in detecting spammer communities. Therefore, spammer community detection based on NSFs is worth studying.

The method proposed in this paper as follows: based on how the reviews were made, the reviewers form a comment-based reviewer network. Thereafter, reviewers can be formed into communities through community detection. We check the structural features of detected spammer communities and find that:

- Different from normal reviewers who tend to have little contact, spammers in spammer communities tightly group together.
- Spammers in the same spammer community behave in a highly similar way.
- Spammer communities have very different network structure compared with normal reviewer communities, due to the different behavior of spammers and normal reviewers. The network structures of spammer communities are characterized by a high clustering coefficient and high self-similarity.

2 Related Work

Early spammers detection mainly concentrated in E-mail [8]. In a wide field, the most investigated spam activities have been in the field of social networks. The first technique in fake review detection was proposed by Jindal et al. [2], who treated duplicate reviews as fake reviews and non-duplicate reviews as truthful reviews. What's more, a rule-based method was proposed [3] to find unusual review patterns by mining unexpected rules as review behavior.

Similarly, behavioral features were used to detect individual spammers. Lim et al. [5] exploited four types of suspicious behavior to detect spammers from the content and rating of a review. Mukherjee et al. [6] treated the concentricity, burstiness and extreme rating of reviews as spammer behavior features, and established a Bayes model to detect spammers. Another approach used the relationship between reviewers [11], calculating the similarity between a spammer and its neighbors, and correcting its label with the votes of neighbors. There also exists a graph-based method [9] computing the trustworthiness of reviewers, the honesty of a review and the reliability of stores, finally creating a ranking list of suspicious reviews and reviewers.

Few studies have focused on spammer community detection. Mukherjee et al. [7] initially proposed the review spammer community detection problem. They used the FIM method to find candidate spammer communities, and ranked the "spamicity" for each community using several behavioral indicators (BIs) derived from the collusion phenomenon of reviewers. The BIs contain the following six indicators: community rating deviation (CRD), community review timestamp gap (CRTG), community reviewed store proportion (CRSP), community size (CS), community early time frame (CETF), and community size ratio (CSR).

Another approach [10] proposed a loose spammer community detection based on the bipartite graph. The method constructed a reviewer network represented as a bipartite graph, and proposed structural indicators (SIs) to evaluate the "spamicity" of detected communities and ranked loose spammer communities. SIs contain the following four indicators: review tightness (RT), neighbor tightness (NT), product tightness (PT), and product reviewer ratio (RR).

3 Methodology

An NSFs-based spammer community detection method is presented in this section. A diagram of the method's architecture is shown in Fig. 1. The first step is creating a comment-based reviewer network based on the review data. The second step is detecting communities in the comment-based reviewer network. The third step is extracting community features of spammer communities. The final step is identifying spammer communities by communities' NSFs.

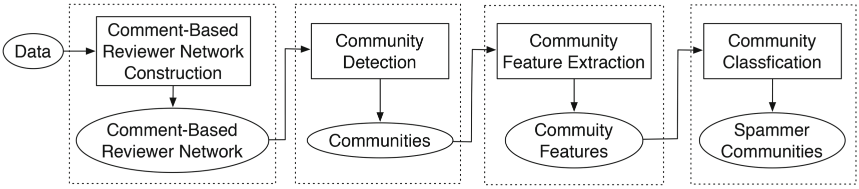

Data
Comment-Based Reviewer Network Construction
Comment-Based Reviewer Network
Community Detection
Communities
Community Feature Extraction
Commuity Features
Community Classfication
Spammer Communities

Fig. 1. The architecture of spammer community detection

3.1 Comment-Based Reviewer Network Construction

The comment-based reviewer network is an undirected weighted graph $G = (V, E, W)$. V is a node set, E is an edge set, $v_i \in V$ is a node, $(v_i, v_j) \in E$ is an edge between v_i and v_j, and each edge has a weight $w_{v_i,v_j} \in [0, 1]$ which presents the similarity between two nodes. Reviewers of online-review sites are mapped onto nodes in the comment-based reviewer network. Each reviewer has a review set $\{u, (s, c, t, r)\}$ including a series of reviews, s is store_id, c is the comprehensive rating of a store, u is the reviewer_id, r is the rating of the review, and t is the post time of the review. For reviewers A and B, the review sets are as follows:

$$R(A) = \{u_A, (s_i, c_i, t_i, r_i)\} \quad i = [1, n];$$
$$R(B) = \{u_B, (s_j^{'}, c_j^{'}, t_j^{'}, r_j^{'})\} \quad j = [1, m].$$

Generally, if two reviewers are neighbors, they have published some collusive reviews [11]. $C_{A,B}(k)$ measures whether a review of A is similar to a review of B. For any two reviewers A and B, $C_{A,B}(k) = 1$, if $(u_A, s_k, c_k, t_k, r_k) \in R(A)$ and $\exists$ $(u_B, s_l^{'}, c_l^{'}, t_l^{'}, r_l^{'}) \in R(B)$, where:

(a) $s_k = s_l^{'}$, which means the two reviews are posted at the same store;
(b) $|t_k - t_l^{'}| < \Delta t$, which means the two reviews are posted within a certain time interval;
(c) $r_k = r_l^{'} = 1$ star or $r_k = r_l^{'} = 5$ stars, which shows the two reviews are posted with an extreme rating.

Different from previous research [10] using Jaccard similarity to measure neighbor tightness between two reviewers, a new similarity measure, Collusion Similarity, is proposed here. The new similarity measure not only depends on the store_id of collusive reviews, but also considers the concentricity and extreme rating of collusive reviews. The collusion similarity between two reviewers A and B is defined as follows:

$$S(A, B) = \frac{\sum_{k=1}^{n} C_{A,B}(k) + \sum_{l=1}^{m} C_{B,A}(l)}{|R(A)| + |R(B)|} \tag{1}$$

where, $|R(A)|$ is the total number of reviews for A. The larger the similarity, the more collusive reviews two reviewers share. If the similarity between two reviewers exceeds the threshold, the pairwise reviewers are considered to be neighbors. The collusion similarity $S(A, B)$ is defined as the weight of edges $w_{A,B}$, so the edge between A and B is $(A, B, w_{A,B})$.

3.2 Community Detection

Community is a structural feature of complex networks. Specifically, the whole network is composed of some "groups" and the internal links of communities are dense, and the links between communities are sparse. Likewise, the comment-based reviewer network consists of some communities.

Despite its role as a normal reviewer or a spammer, a node only belongs to one community. Therefore, non-overlapping community detection can perform well. We use a fast unfolding method to detect communities on large undirected weighted networks, as shown below:

Algorithm 1. Community detection on undirected weighted networks

```
Input: the comment-based reviewer network $G = (V, E, W)$
Output: communities, optimal modularity $Q$
initialize each node as a community;
calculate the modularity $Q$ of the initial network;
repeat
    calculate the number of communities $n$;
    for each $i \in [1, n]$ do
        calculate the number of neighbors $m$ for $i$;
        for each $j \in [1, m]$ do
            calculate the change of modularity $\Delta Q$ if assign $i$ to a the community where $j$ is located;
        end for
        assign $i$ to the community with $j$ where $\Delta Q\ ! = 0$ and $\Delta Q$ is highest
    end for
    compress $G$ into $G^{'}$: compressing one community into one new node; compressing the weights of all links between two communities into one weight of the link between two new nodes; and compressing the weights of all links within the community into one weight of the ring of a new node;
    calculate the modularity $Q$ of $G^{'}$;
until the modularity $Q$ no longer changes;
```

3.3 Network Structural Features

Most previous research has put forward a series of behavioral indicators to detect spammer communities. These indicators, however, do not apply to the current situation. Further, certain structural indicators were used to improve the effect of classification, but their performance was found to be poor [1].

However, the network structure of spammer communities provides a novel idea for spammer community detection. NSFs can identify some spammer communities, in which the behavior of spammers tends to normal reviewers. As shown in Fig. 2, the links between reviewers of a normal reviewer community are sparse, but the links between reviewers of a spammer community are dense. This example suggests that the network structure within the normal reviewer community and the spammer community is different. Therefore, NSFs can be used as an important basis for spammer community detection. The NSFs of a community are defined in the following:

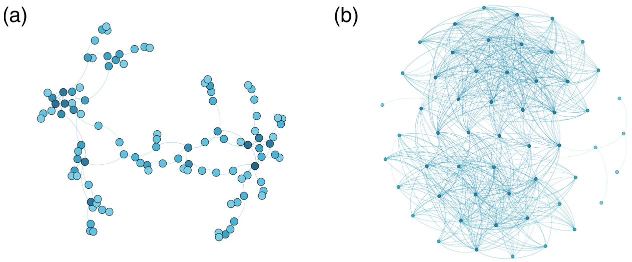


Fig. 2. Comparison between normal community an spammer community. (a) An example of normal community. (b) An example of spammer community.

Self-Similarity (SS).

$$SS = \frac{\sum_{(v_i,v_j)\in E} w_{v_i,v_j}}{|E|} \tag{2}$$

where $|E|$ is the number of edges in a community and w_{v_i,v_j} is the similarity between pairwise nodes v_i and v_j. A spammer community is an organized group with high self-similarity. In order to efficiently complete tasks, spammers post more collusive reviews, the similarity between two spammers is extremely high. And, the local structure is similar with the global structure.

Clustering Coefficient (CC).

$$CC = \frac{\sum_{v_i\in V} \frac{2n}{k(k-1)}}{|V|} \tag{3}$$

where $|V|$ is the number of community members, k is the degree of v_i, n is the number of edges between the neighbors of v_i. The clustering coefficient reflects the tightness of community structure. Most pairwise members in a spammer community post collusive reviews, the link density of nodes in the spammer community is very large. The higher the clustering coefficient, the greater the suspiciousness of that community.

To quantify the importance of network structure in detecting spammer communities, this paper studies a series of samples (a manual annotated data set from Dianping) including spammer communities and normal reviewer communities. Analysis of these samples in the comment-based reviewer network reveals three results (As shown in Fig. 3):

- For fixed BIs or SIs, there is a wide difference between the state of samples. The larger the CC of a sample, the greater the probability that the sample is a spammer community. In particular, there are many samples (large BIs, low CC or large SIs, low CC) that are normal reviewer communities.

- For a fixed CC, the state of a sample is approximately independent of the BIs and SIs. This result is revealed in the vertical structure, suggesting that whether a sample is a spammer community depends on the CC of the sample, and samples with large CC are spammer communities.
- Similarly, the SS of a sample determines whether it is a spammer community more than BIs and SIs. But, the performance of SS is worse than that of CC.

In summary, the above results illustrate the fact that suspicion is larger for communities of higher CC or SS, whereas communities of a given BIs or SIs value can result in either small or large suspicion, depending on the value of NSFs. Therefore, spammer communities in the comment-based reviewer network are not necessarily related to the BIs and the SIs. Instead, NSFs are a better predictor of spammer communities and CC has the best performance.

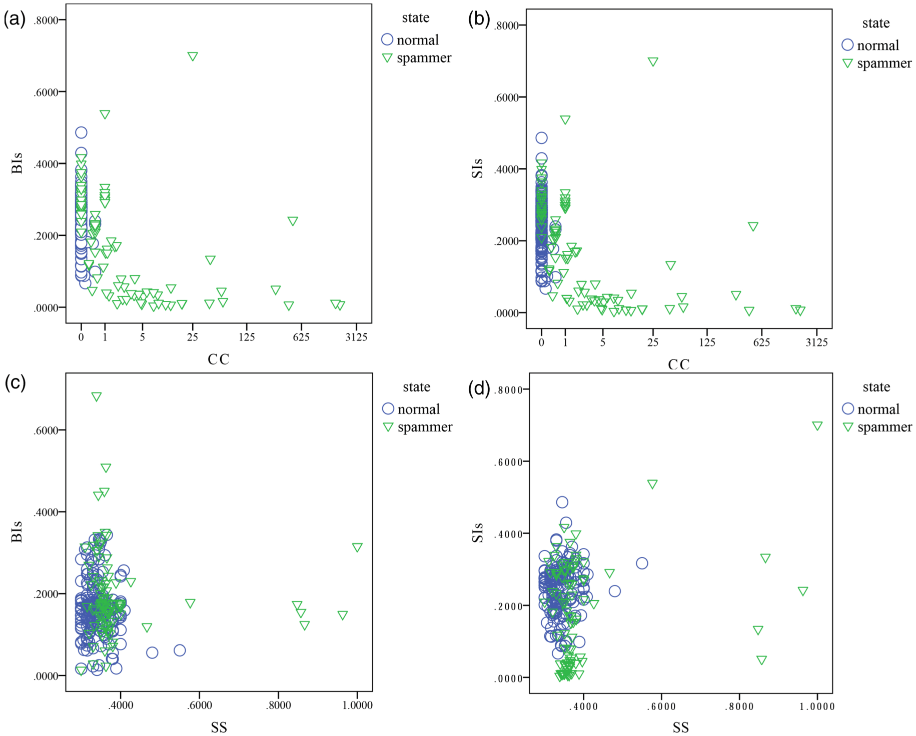
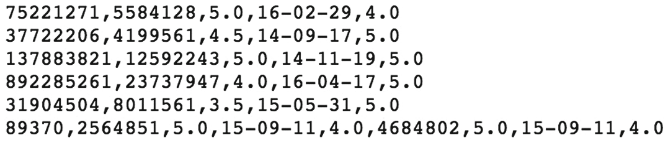


Fig. 3. Analysis of samples from different dimensions. (a) The CC of samples is compared with the BIs of samples. (b) The CC of samples is compared with the SIs of samples. (c) The SS of samples is compared with the BIs of samples. (d) The SS of samples is compared with the SIs of samples. The circles denote the samples of normal reviewer communities; the triangles denote the samples of spammer communities.

4 Experiment and Analysis

4.1 Dataset and Community Detection

In this section, we apply spammer community detection through NSFs to a large-scale data set including 5,427 stores, 1,669,060 reviewers and 2,920,122 reviews collected from January 2014 to December 2016 from the online-review website Dianping (https://raw.githubusercontent.com/SaraLiu1994/Dianping-Dataset/master/review.txt). The sample of dataset is shown in Fig. 4. Each line is a review set of one reviewer. A review set contains one or more reviews published by the reviewer. For example, the last reviewer in Fig. 4 has two reviews and other reviewers have only one review respectively. In a line, the first field is reviewer_ id. The rest are its reviews. Each review contains 4 fields including s - store id, c - the comprehensive rating of s, t - the post time of the review, and r - the rating of the review.



```
75221271,5584128,5.0,16-02-29,4.0
37722206,4199561,4.5,14-09-17,5.0
137883821,12592243,5.0,14-11-19,5.0
892285261,23737947,4.0,16-04-17,5.0
31904504,8011561,3.5,15-05-31,5.0
89370,2564851,5.0,15-09-11,4.0,4684802,5.0,15-09-11,4.0
```

Fig. 4. Extract from main dataset.

In order to reduce the impact of noise, some reviewers who write few reviews are removed. The final review set contains 89,006 reviewers and 709,220 reviews. Thereafter, we create the comment-based reviewer network by using the collusion similarity measure described in Sect. 3.1. Next, we divide the comment-based reviewer network into communities by using a non-overlapping community detection algorithm. At this point, the optimal modularity is 0.806, and the comment-based reviewer network consists of 289 communities (communities in which the number of members is less than three are removed). After manual evaluation, 289 communities are labeled into 89 spammers communities and 200 normal reviewer communities. The training set contains 50 spammer communities and 105 normal reviewer communities, and the test set contains 39 spammer communities and 95 normal reviewer communities.

4.2 Evaluation

Spammer community classification is a binary classification problem. The evaluation indexes of binary classification include precision, recall, F1 and AUC. Precision refers to how many real positive samples in all positive samples are predicted by the classifier. Recall refers to how many real positive samples are predicted as positive samples. F1 can be regarded as a weighted average of precision and recall, used when precision clashes with recall. AUC is the area under the ROC curve commonly being used to measure classification performance. The larger the AUC, the better the classification performance.

As shown in Table 1, we compare the classification performance of each indicator using different evaluation indices. We find:

- In terms of precision, CC performs the best, CRD preforms the worst which reveals that spammers reduce the deviation by publishing little truthful reviews.
- In terms of recall, CC performs the best, CETF preforms the worst which confirms that the time interval of fake reviews is not concentrated in the early days.
- In terms of F1, CC performs the best, CETF preforms the worst.
- In terms of AUC, CRD, CETF and CSR perform poorly, which proves spammer communities weaken their behavioral features to hide their true identity. But, CC does best with 0.969 AUC which confirms that it is a good indicator for spammer community detection.
- Although NSFs contain only two indicators (SS, CC) which is fewer than SIs (PT, NT, RT, RR), the NSFs perform the best with 83.72% precision, 92.31% recall, 87.80% F1 and 0.959 AUC.

Table 1. The performance of each indicator

	SS	CC	CRD	CRTG	CRSP	CS	CETF	CSR	NT	PT	RT	RR	NSFs	BIs	SIs
P(%)	71.88	**94.59**	21.43	45.71	32.88	29.33	28.57	35.48	80.00	43.38	63.16	73.91	**83.72**	37.93	72.41
R(%)	58.97	**89.74**	15.38	41.03	61.54	56.41	5.13	28.21	51.28	76.92	30.77	43.59	**92.31**	28.21	53.85
F1(%)	64.79	**92.11**	17.91	43.24	42.86	38.60	8.70	31.43	62.50	55.56	41.38	54.84	**87.80**	32.35	61.76
AUC	0.692	**0.969**	0.441	0.605	0.652	0.632	0.500	0.544	0.612	0.698	0.619	0.652	**0.959**	0.599	0.735

The above results illustrate the fact that SIs can improve the accuracy of spammer community detection, but have limited effectiveness. Thus, spammer community detection is not necessarily related to the BIs or SIs. Instead, NSFs can contribute to spammer community detection, with CC being the best predictor of spammer communities.

5 Conclusion

In this paper, reviewers form a comment-based reviewer network based on their reviews. Reviewers can be formed into communities through community detection and we check the structural characteristics of spammer communities. Our findings include: (a) the behavior of spammers has become more hidden and it is more difficult to identify spammers using only BIs or SIs. (b) Compared with normal reviewer communities, spammer communities have different network structural features, a high clustering coefficient and high self-similarity. (c) Compared with BIs, SIs can improve the accuracy of spammer community detection, but their effectiveness is limited. Instead, CC is the best predictor of spammer communities.

This work aims to use NSFs to identify spammer communities in comment-based reviewer networks, and has achieved good performance on real-world review data. However, many different types of community structure have been proposed. Whether these community structure types can further distinguish spammer communities from normal reviewer communities should be evaluated in future research.

Acknowledgments. This research is supported by the National Natural Science Foundation of China (Grant No. 71203135). Thank Nikita Koptyug for proofreading our manuscript.

References

1. Heydari, A., ali Tavakoli, M., Salim, N., Heydari, Z.: Detection of review spam: a survey. Expert Syst. Appl. **42**, 3634–3642 (2015)
2. Jindal, N., Liu, B.: Opinion spam and analysis. In: Proceedings of the 2008 International Conference on Web Search and Data Mining, pp. 219–230. ACM
3. Jindal, N., Liu, B., Lim, E.P.: Finding unusual review patterns using unexpected rules. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, pp. 1549–1552. ACM (2010)
4. Li, H., Liu, B., Mukherjee, A., Shao, J.: Spotting fake reviews using positive-unlabeled learning. Computación y Sistemas **18**, 467–475 (2014)
5. Lim, E.P., Nguyen, V.A., Jindal, N., Liu, B., Lauw, H.W.: Detecting product review spammers using rating behaviors. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, pp. 939–948. ACM (2010)
6. Mukherjee, A., Kumar, A., Liu, B., Wang, J., Hsu, M., Castellanos, M., Ghosh, R.: Spotting opinion spammers using behavioral footprints. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 632–640. ACM (2013)
7. Mukherjee, A., Liu, B., Glance, N.: Spotting fake reviewer groups in consumer reviews. In: Proceedings of the 21st International Conference on World Wide Web, pp. 191–200. ACM (2012)
8. Wang, D., Irani, D., Pu, C.: A study on evolution of email spam over fifteen years. In: International Conference on Collaborative Computing: Networking, Applications and Worksharing, pp. 1–10. IEEE
9. Wang, G., Xie, S., Liu, B., Philip, S.Y.: Review graph based online store review spammer detection. In: 2011 11th IEEE International Conference on Data Mining, pp. 1242–1247. IEEE (2011)
10. Wang, Z., Hou, T., Song, D., Li, Z., Kong, T.: Detecting review spammer groups via bipartite graph projection. Comput. J. **59**, 861–874 (2016)
11. Xu, C., Zhang, J., Chang, K., Long, C.: Uncovering collusive spammers in Chinese review websites. In: Proceedings of the 22nd ACM International Conference on Information and Knowledge Management, pp. 979–988. ACM (2013)
12. Xu, Y., Li, F., Liu, J., Zhang, R., Yao, Y., Zhang, D.: Detecting false information of social network in big data. In: Wang, S., Zhou, A. (eds.) CollaborateCom 2016. LNICST, vol. 201, pp. 642–651. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59288-6_65
13. Zhou, T., Bai, W., Wang, B., Liu, Z., Yan, G.: A brief review of complex networks. Physics **34**, 31–36 (2005)