

INTERNET CORE AND PERIPHERY FROM A MARKETING COMMUNICATIONS PERSPECTIVE

Abstract. In this article, traditional approaches to Internet metrics are upgraded as part of a lengthy study with insights reached by utilising network analysis methods to improve understanding of conventional reach figures. Groups of structurally equivalent sites and browsing probabilities show relations between websites' audiences and provide information missing in reach figures. A synergetic interpretation of both types of information enables more sophisticated marketing communications planning in the Internet. A national network of competing commercial sites is analysed as a case study.

Keywords: Internet Marketing Communications, Websites' Reach, Exploratory Network Analysis Methods, Pajek, Network Visualisation, Blockmodeling, Structural Equivalence, Probabilistic Link Weights, Browsing Probabilities

Introduction

In the article, we summarise and supplement our previous research on relations between audiences of websites to provide a final and more comprehensive case of combining reach figures with exploratory network analysis outcomes. The stress is not on a theoretical framework and basic concepts, as in Cucin et al. (2008), or on analytical procedures, as in Kropivnik and Kejžar (2011), but more on the art of synergic interpretations of outcomes from a marketing communications and particularly from an advertising perspective.

Following the main line of our lengthy research we first provide a brief summary of arguments for the necessity of supplementing a website's reach figures with information about relations between audiences, then present the methodological basis of obtaining missing information, and finally

* Samo Kropivnik, PhD, Associate Professor in social sciences research methodology, Faculty of Social Sciences, University of Ljubljana; Nataša Kejžar, researcher, Institute of Biomedical Informatics, Faculty of Medicine, University of Ljubljana.

focus on combining reach figures with both structural positions in the network and probabilistic links between websites in a productive way.

The advantages and the drawbacks of a website's reach figures

The reach is a universally accepted currency in the advertising industry (Belch and Belch, 2004) that can also be adapted to Internet advertising. After agreement on a definition is achieved, the reach can be utilised both in traditional and in new media, which makes reach figures so broadly accepted and popular among marketing communicators, media experts and researchers. However, as we are about to demonstrate, in the case of the Internet additional insight can be provided to make the best use of reach figures.

Slovene Internet advertising is explored as a case study, since it is characterised by some useful features: in terms of language, the Slovene share of the Internet is relatively small (with a limited number of sites, publishers and users) and is therefore easier to become fully acquainted with, is well developed in terms of technical features (e.g. connections, accessibility), in terms of the proportion of users in the population, in richness of content, and last but not least very well measured (the Internet audience measurement project MOSS has been running since 2005).

In Slovenia, a mixed research model is used in order to determine reach figures. The model incorporates an automatic measurement, a telephone survey and a web survey. The automatic measurement quantifies the activity of computers, while the telephone and the web survey are used to convert figures about computers uploading sites into individuals visiting sites as well as to add demographic, attitudinal and behavioural profiles of users. As a result, the reach refers to the number of different individuals who have visited each website in the period of measurement.

In autumn 2006, all significant national commercial sites but one took part in the measurement. Their reach figures ranged from 2,210 to 710,106 users within the four-week period. The reach of the great majority of websites was lower than 50,000 individuals (i.e. roughly 5% of the online population). One sixth of the sites had a reach higher than 100,000 (i.e. approximately 10% of the online population), and only a few higher than 200,000. There was only one site with conspicuous reach, providing services to three quarters of the online population (Cucin et al., 2008). Such a distribution is not unexpected: in a national context one can expect a relatively small number of highly popular sites and a large number of far less visited sites, although the ratio between them and the reach diversity can vary. Nevertheless, it would be very unusual to find all sites equally popular or equally unpopular.

It should be stressed that what is meant by an Internet site is by no means a single page, but rather a cluster of Internet pages, belonging to the same professional Slovene Internet publisher and sharing similar, organised and interconnected content. Typical examples are Slovene search portals, national and commercial TV stations' sites, news portals, automotive sites and various specialised online services portals.

The reach figures presented above are the main and most utilised result of the MOSS project. The figures demonstrate the capacity of each site in terms of its coverage of the Slovenian Internet population. These are definitely basic and valuable statistics for marketing communications strategies, which should never be ignored in advertising planning. On the other hand, it is obvious that these figures give no insight into relations between audiences of Internet sites, which is a serious problem in their applications. Since individuals may shift from website to website while browsing the Internet, audiences fluctuate much more than was possible in traditional media. Hypothetically, in the core the same users could visit all the internet sites or each site could be visited by different users. Audiences could be perfectly related, meaning that smaller audiences of less popular sites are only subgroups of larger audiences of more popular sites. In general, therefore, only the most popular website with the highest reach really matters for advertising and all other websites are practically unimportant. Alternatively, audiences could be completely independent, meaning that each website covers a different segment of the online population. If so, regardless of the reach, each website matters for advertising (none can be excluded) and campaigns must be carefully planned to reach targeted segments.

At this point an additional insight is required to improve the understanding and the use of reach figures. Two approaches are conventional in marketing communications practice. The more formal solution is to recalculate the reach statistics for the selected sites in such a way as to identify the audience overlap between them and gain the unique reach (Cohen, 1993). However, even a relatively small number of sites and relatively small online populations can demand an unreasonable amount of resources and produce an unmanageable amount of results, since the figures have to be recalculated for all possible combinations of sites – even our case of quite a low number of websites (namely 83) would mean almost 10^{25} different combinations. The more intuitive solution is to rely on the content of webpages and/or on the socio-demographic profiles of users to estimate the overlap between audiences (e.g. about zero overlap between sites with incompatible content, such as children's cartoons and stock market trends, or between sites visited more by youngsters and sites visited more by older people. In both cases, however, such estimations are nothing more than educated guesses which, methodologically speaking, cannot be validated: assumptions about

incompatibility, even if perfectly logical, can be wrong because of the multiple identities of modern consumers and the fact that conclusions about individuals simply cannot be drawn from aggregated data.

What we suggest is a third, far more systematic and formal, method. We suggest the adoption of a more analytical perspective and construct a picture (or a model) revealing how people browse the Internet. The required model should include currently known information (i.e. the reach figures) and upgrade them with the insight missed until now (i.e. information revealing relations between audiences). Such a picture should improve the understanding of reach figures by disclosing actual relations between the audiences of Internet sites and so enable more sophisticated marketing communications planning.

The exploratory network analysis approach

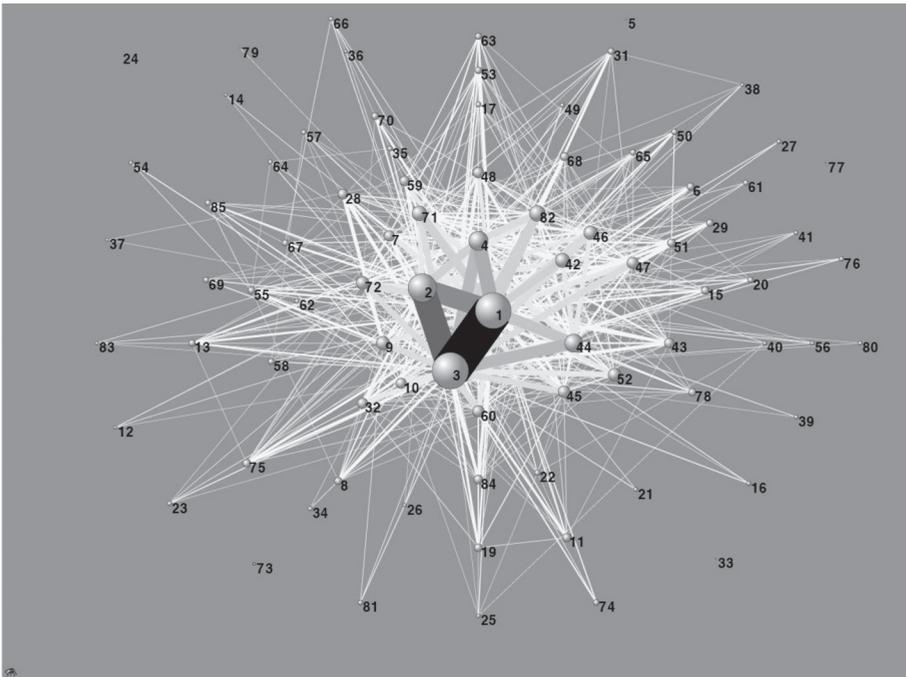
Since we are dealing with a problem of simultaneously analysing attributes of Internet sites (the reach figures) and relations between Internet sites (the audiences' overlap), network analysis methods have been recognised as being particularly well-suited to the problem. From the point of view of the descriptive network analysis approach (see Wasserman and Faust, 1994 or Newman, 2008), the Slovenian Internet can be expressed in terms of patterns or regularities between the interacting sites. Such patterns can be effectively visualised and presented in a graph or described with key statistics. What is specific in our approach is that these patterns depend completely on the browsing behaviour of online audiences and not on publishers' (or owners', or authors') hypertext links embedded in the sites. Conveniently, the data are already available since in automatic Internet traffic measurement dimensions that are not articulated in the reach figures but that can be expressed in network analysis results are present.

As in previous research, we used a sample of 2328 Internet users from 2006 (for data details see Cucin et al., 2008) to construct a network of users visiting websites. A network composed of vertices (websites) and links between vertices (relations between audiences) has been created and visualised by a network analysis computer program package Pajek (de Nooy et al., 2005 and Pajek, 2008). Additionally, we have separated regular links (i.e. relevant, systematic links occurring because visiting both sites is a common pattern, for example) from arbitrary links (i.e. irrelevant unimportant links occurring by chance, because the other site was visited by chance, for example, or because of a unique cause) and visualise the network on the basis of only regular links in order to obtain a more realistic picture of relations between websites (for details see Kropivnik and Kejžar, 2011).

As the key information, the size of vertices in the following picture is

presented relative to their audience (i.e. to the reach). Sites with a high number of visitors are presented with large circles and are easy to identify. The thickness and darkness of links represent their weight: the weight of a certain link depends on how many individuals have visited both websites. “Must-see” websites have links with high weights, because people always stop to visit them while browsing through the Internet, and are therefore also easy to see.

Figure 1: VISUALISATION OF THE NETWORK



As stated above in the comments on the reach figures in the previous section, there are only a few sites with a significantly high reach. What now becomes clear is that the very same sites are connected with high weight links, meaning that they have highly related audiences. Advertising on these websites guarantees a higher chance of being seen (because of the high reach) but at the same time increases the possibility of hitting the same users with the same advertisements over and over again (because of the high overlap of audiences). The high frequency of advertising on these websites (a common pattern), risks not only a loss of financial efficiency but also being perceived as annoying and a decline in potential consumers. To avoid this, both pieces of information (i.e. the reach and the weight of links) should be considered jointly.

Furthermore, links between less frequently sites help us to find out which subsegments of the Internet population are independent (no links between sites), which are highly related (high weight links) and to what extent small audiences of specific sites are related to large audiences of the most visited sites.

How all these pieces of information are exactly combined depends on the specifics of marketing communication campaigns and cannot be universally prescribed. Generally, however, with better insight the total effectiveness of campaigns can be increased without raising costs and, even more importantly, without raising the risk of treating the online population indiscriminately by always feeding them with the same advertisements if Internet marketing communications are planned according to websites' positions in the network not just according to their content and the size of the audience.

Structurally equivalent sites and browsing probabilities¹

In order to recognise groups of structurally equivalent sites (i.e. sites for which the same principles can be applied in marketing communications planning due to their similar positions in the network) we have applied the sum of squares homogeneity blockmodeling approach (Doreian et al., 2005; Žiberna, 2007). This approach draws on structural equivalences between vertices based on the homogeneity of blocks. Structurally equivalent vertices are vertices that have exactly the same links to other vertices (Lorrain and White, 1971), and blocks are optimised in order to homogenise the weight of links inside each block as much as possible.

Generally speaking, the vertices inside the same block (group) could be interchanged since they are structurally equivalent: in other words, they are almost the same in the way they are connected to vertices in the same block, and simultaneously almost the same in how they are connected to vertices in other blocks. Translated into a description of a network of websites, pairs of sites in a group share approximately the same number of mutual visitors and, likewise, each site in a group shares approximately the same number of mutual visitors with sites in other groups.

In fact, each group of Internet sites is characterised by the strength of inter-group relations and by relations with other groups. To stratify the whole network, the absolute strength of the links is required, because frequent streams of visitors had to be distinguished from rare ones, but for a more comprehensive insight different relations' aspects are required in addition to absolute strength. Therefore, probabilistic link weights as

¹ *All procedures regarding homogeneity block modelling and applying probability links are more exhaustively presented in Kropivnik and Kejžar, 2011.*

average degrees of overlap between audiences of different sites are calculated and applied to conclude the analysis. The weight and the reach values of the selected neighbouring link and vertex are taken and divided. To obtain the probability for the link of the opposite direction the reach of the end-vertex of the same link is taken into account. The browsing probabilities tell us how likely it is for an average user who visits a particular site to visit each other site in the network.

Figure 2: RESULTS OF SUM OF SQUARES HOMOGENEITY BLOCKMODELING INTO FOUR GROUPS

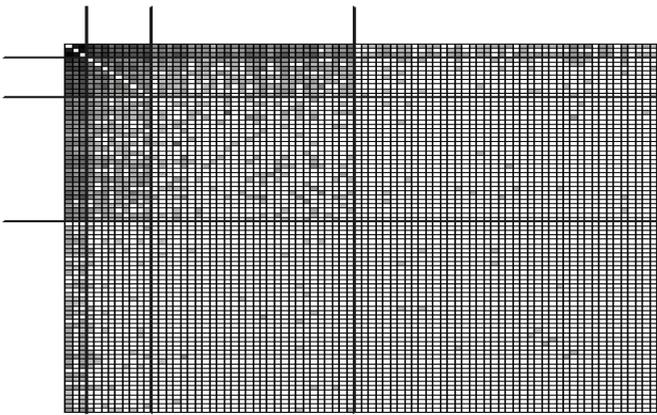
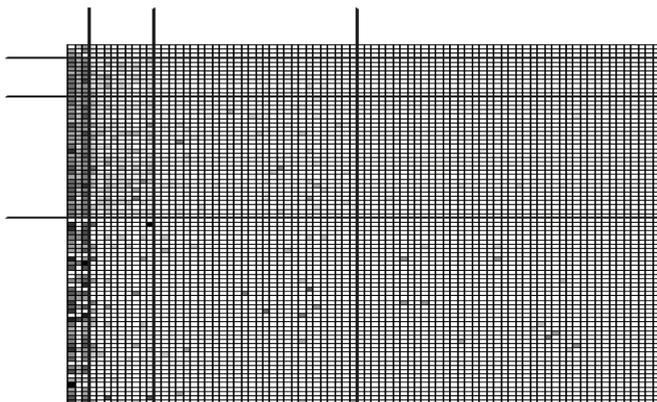


Figure 3: BROWSING PROBABILITIES PROJECTED ON THE BLOCKMODELING RESULTS



The final results are presented below in two figures that have to be considered together. Figure 2 shows the result of homogeneity blockmodeling in four groups. The network is presented as a symmetric matrix: each row A (just as each column B) represents a vertex (a website) and the intensity of its links (relations, number of shared visitors) to other vertices – the darker the cell the larger the weight of that link (the higher the number of mutual visitors). In Figure 3, browsing probabilities are projected into the matrix: the darkness of the cell AB represents the probability that an average user who visits site A also visits site B. The darker the square the higher the probability (the higher the overlap between audiences).²

The core and the periphery

Figure 2 clearly suggests that the Slovene Internet network can be characterised as a “core-periphery structure”, composed of a small number of highly interconnected sites, some sites that are somehow connected in the network and a large number of very weakly connected vertices with only a few links, all of low weights. The hierarchical structure can be clearly recognised and confirmed since dark squares quickly disappear if we move from the left upper corner to the right bottom corner of the matrix or simply from the top to the bottom (alternatively from left to right). Such a pattern is considered in network analysis as a core-periphery structure with deferential ties between periphery levels (see Wasserman and Faust, 1994 and White et al., 1976). There are tree vertices in the highly interconnected and connected-to-others core (the first block), and all the subsequent blocks are more and more peripheral, meaning that they are still strongly connected to the core, less to the blocks above them (i.e. the more connected clusters) and far less to the least connected group overall. This has to be taken into account in planning marketing communications in the Internet.

Additionally, the relative degree of overlap between audiences of different sites is visible from Figure 3, and these valuable pieces of information improve hierarchical core-periphery block descriptions. Finally, the reach figures have to be reconsidered in combination with the sites’ positions in the network. Only a consideration of both types of parameters taken together enables us to recognise and utilise the unique character of each of the four groups of websites, as described below.

² *Actual probabilities could be printed in the cells, but would be visible only in large prints.*

The “must-see” websites

The first group of websites is the smallest, composed of only three sites. These are the sites with the highest reach figures and consequently commonly used in marketing communications. Sites from the first group are characterised both by a large number of mutually shared visitors and quite a large number of visitors shared with other sites - more with the second-group sites, less with the third-group sites and far less with the fourth-group sites (Figure 2). It can be speculated that all three are the must-see sites which users are almost compelled to visit each time they browse the Internet. While they are thus ideal in reaching dispersed users of the great majority of other sites, it is potentially dangerous to satiate users with advertisements if all first-group sites are included in the same campaign at the same time. However, browsing probabilities (Figure 3) show them to be less inter-connected than perceived from absolute link weights. Due to the high reach figures in this cluster, very high link weights turn out to be at most medium high probabilities. An average user that visited one of the sites is just about 37% likely (probabilities of 0.23, 0.44, 0.46, 0.36, 0.43, 0.28) to also visit any of the other two sites in this cluster (in Figure 3, the cells in the top left block are not very dark). Therefore, the danger of satiating users with advertisements if all first group sites are included in the campaign is only modest, although it cannot be neglected. At the same time, their potential to reach dispersed users of the great majority of other sites is even bigger, since there are high probabilities of sharing audiences with most of the other sites (in Figure 3, the cells in the first three columns are the darkest ones). For an average user that visited one of the other sites (from the other three groups) a visit to one of the most popular sites is quite likely. For most of the sites, likelihoods are between 30% and 60%, for some even up to 100% and below 30% only for a few (including flat 0%).

When sites from the first cluster are further considered on an individual basis, it becomes particularly interesting that only two of them highly attract visitors from other sites (see the first and the third column in the matrix, Figure 3). If we consider only browsing probabilities higher than two-thirds, the first site attracts visitors from fourteen other sites, the second from only one site from other clusters, and the third from eighteen other sites. Additionally, there is almost no overlap between the list of sites significantly oriented towards the first site and the list of sites significantly oriented towards the third site and the overlap between the audiences of these two sites is estimated at about 40% (approximately the same in both directions). It is no secret that these two “winning” sites are a popular local search engine and the commercial TV station site, both already playing an important part in Internet advertising.

The “quite popular” websites

This group is the second smallest, composed of nine sites with quite different, but as a rule, high reach figures and this also makes them tempting for advertising. These sites are more difficult to describe, since they are characterised by quite a large number of mutually shared visitors, a large number of visitors shared with the first-group sites (overlap with “must-see” sites), still considerable overlap with the third-group sites and almost no shared visitors with the fourth-group sites (Figure 2). It can be speculated that they are quite popular sites that a lot of users often visit when they browse the Internet. They seem to be still effective in reaching dispersed users of the majority of other sites, with the notable exception of the fourth-group sites, and less liable to satiate users with advertisements if the whole group is included in the campaign. Browsing probabilities (Figure 3) show them to be less inter-connected than sites from the first group (the likelihoods are lower, mostly between 10% and 20%), so the danger of satiating users is also correspondingly lower, almost negligible in fact. At the same time, all other probabilities are in general lower, meaning that an average user of other sites is considerably less likely to visit sites in this cluster than sites in the first cluster. The second group of websites is thus in fact not particularly effective in reaching dispersed users of the Internet. However, there are a few noticeable exceptions (see darker squares in the third and fourth block) to remind us not to forget about the individuality of each site. In marketing communications planning, as well as being considered as a group these websites have to be approached individually.

The “specialised” sites

This group is the second largest, composed of twenty-eight sites with very diverse reach figures from almost the top to almost the bottom. Sites from the third group are mutually quite unrelated: generally they do not share a high number of visitors, with a few pairs of evident exceptions (see the darker cells in Figure 2). At the same time, they share a considerable number of visitors with the first two groups (as already stated above), but almost none with the fourth group of sites. It can be speculated that the sites are quite specialised and the audience of this group quite segmented according to their desires (interests, hobbies, brands etc.), but still in the mainstream in terms of more popular sites, which they often visit. With regard to browsing probabilities, we find considerably lower, almost negligible, probabilities of sharing audiences inside the group and with the fourth group, and modest probabilities of sharing audiences with the first two groups. Again, however, there are some quite conspicuous exceptions

(see the darker squares in Figure 3) that prevent deterministic conclusions. In marketing communications planning the third group audiences can be partially reached through popular sites; however, for more inclusive audience coverage sites from the third group should be considered more or less individually.

The “isolated” websites

Notably, this is the biggest group, composed of forty-three sites with generally the lowest reach figures (again, with conspicuous exceptions). The websites from the last group are unrelated to each other and simultaneously to all other sites but the most popular three websites (the first group, to which they are only modestly related). In general, the almost independent audiences cannot be reached through more popular sites (Figures 2 and 3). It can be speculated that they are isolated, specialised sites, fulfilling uncorrelated desires. In marketing communications planning these sites have to be considered exclusively on an individual basis, to address narrowly-defined target groups, drawing mostly on their focused content and less on the reach and browsing probabilities.

Conclusions

As demonstrated, interpretations of reach figures can be improved with an insight into relations between audiences of Internet sites to enable more sophisticated marketing communications planning in the mysterious and unpredictable world of the Internet. The reach provides a base for a site's position but it cannot be taken as the only factor in determining a site's role in the network. The reach figures demonstrate the potentials, not the actual realisations of these potentials, which are embodied in a sites position and role in the network.

To sum up the results, the smaller two of the four groups of websites are, in general, more embedded in the network and the larger two are, roughly speaking, more isolated. It seems that most of the sites with high reach manage to turn it into an advantage, but not all of them. Most sites with a high reach are in the first two groups and some belong to the third group. On the other hand, some sites with a lower reach can be found outside the fourth group. When the groups are considered separately, the first definitely comes out as the most tempting for advertising as the type of sites with the highest reach figures, modestly-shared audiences inside a block and large flows of visitors from the great majority of other sites. To a certain extent, the second group also shares these virtues. Nevertheless, in both groups some sites are more suitable for advertising and others less so. The level of suitability can

be recognised as a combination of a high reach, a low audience overlap with other sites included in the campaign and a high overlap of audiences with sites that are not included in the campaign.

The study demonstrates the capacities of exploratory network analysis in marketing communications planning in the Internet, although certain limitations exist. Firstly, only a sample of 2328 Internet users has been analysed – some audiences are therefore quite small and consequently some links, patterns and probabilities are based on small groups of people. Given estimations, especially probabilities, can be quite rough and should be generalised with caution. The solution is to enlarge the sample or use the whole population in further studies to obtain more precise estimations.

Secondly, only a national segment of the Internet is analysed, although the Internet is a global media and its users do not stay inside virtual national or linguistic borders. Nevertheless, despite the globalisation of communications, sites communicating in national languages represent a base for successful marketing communications.

While the obtained results can be questioned due to the limitations of the study stated above, the potential of the methodological approach can nevertheless be recognised. We believe that exploratory network analysis has proven to be more than useful in providing additional insight into the fundamental metrics of Internet marketing communications planning.

REFERENCES

- Aho, Alfred V., Michael R. Garey and Jeffrey D. Ullman (1972): The Transitive Reduction of a Directed Graph. *SIAM Journal on Computing*, 1(2): 131–137.
- Belch, George E. and Michel A. Belch (2004): Advertising and promotion: an integrated marketing communication perspective. New York: McGraw-Hill/Irwin.
- Breiger, Ronald L. (1974): The duality of persons and groups. *Social Forces*, 53: 181–190.
- Chaffey, Dave, Richard Mayer, Kevin Johnston and Fiona Ellis-Chadwick (2006): *Internet Marketing: Strategy, Implementation and Practice*. Harlow: Financial Times/Prentice Hall.
- Coupey, Eloise (2001): *Marketing and the Internet*. New Jersey: Prentice Hall.
- Cucin, Patricia, Nataša Kejžar and Samo Kropivnik (2008): Internet advertising planning with the network approach. V Klement Podnar in Zlatko Jančič (ur.), *Corporate and marketing communications as a strategic resource*, 13th International Conference on Corporate and Marketing Communications – CMC, 196–202. Ljubljana: Routledge.
- Doreian, Patric, Vlado Batagelj and Anuška Ferligoj (2005): *Generalized Blockmodeling*. New York: Cambridge University Press.
- Grönroos, Christian (1989): Defining Marketing: A market oriented approach. *European Journal of Marketing*, 23 (1): 52–60.

- Hagel, John and Arthur G. Armstrong (1997): *Net Gain: expanding markets through virtual communities*. Boston: Harvard business School Press Boston.
- Kropivnik, Samo and Nataša Kejžar (2011): Potenciali opisne analize omrežja za trženjsko načrtovanje na internetu. *Teorija in praksa*, 48(1): 45–69
- Lorrain, F. and H.C. White (1971): Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology*, 1 (January): 49–80.
- Newman, M. E. J. (2008): Mathematics of networks. V Blume, L. E. in Durlauf, S.N. (ur.), *The New Palgrave Encyclopedia of Economics*. Basingstoke: Palgrave Macmillan.
- de Nooy, Wouter, Andrej Mrvar and Vladimir Batagelj (2005): *Exploratory Social Network Analysis with Pajek*. New York, NY: Cambridge University Press.
- Ordahl Thomas (2004): The Problem with Internet Advertising. Dostopno preko <http://www.imediaconnection.com/content/2905.asp>, 16. 2. 2008.
- Pajek, Program for Large Network Analysis (2008). Dostopno preko <http://pajek.imfm.si>, 9. 12. 2008.
- Landry, Edvard Carolyn Ude and Christopher Vollmer (2007): *HD Marketing 2010: Sharpening the Conversation* (white paper). Dostopno preko http://www.iab.net/insights_research/iab_news_article/64401?o12499, 6. 2. 2008.
- Strauss, Judy, Adel El-Ansary and Raymond Frost (2006): *E-Marketing*. New Jersey: Prentice Hall.
- Žiberna, Aleš (2007): Generalized blockmodeling of valued networks. *Social Networks*, 29 (1): 105–126.
- Ward, Joe H. Jr. (1963): Hierarchical grouping to optimize an objective function. *Journal of American Statistical Association*, 58 (301): 236–244.
- Wasserman, Stanley and Katherine Faust (1994): *Social Network Analysis*. Cambridge University Press.
- White, Harrison C., Scott A. Boorman and Ronald L. Breiger (1976): Social structure from multiple networks I. Blockmodels of roles and positions, *American Journal of Sociology*, 81: 730–779.