Krzysztof Janc

Possibilities of hyperlink application in spatial research

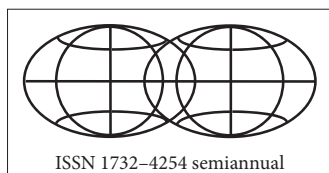MUZEUM HISTORII POLSKI

Possibilities of hyperlink application in spatial research

Krzysztof Janc

University of Wrocław, Institute of Geography and Regional Development, Kuźnicza 49/55, 50-138 Wrocław, Poland;
phone: +48 713 752 246, e-mail: krzysztof.janc@uni.wroc.pl

Abstract. The main aim of the paper is to show the selected ways of analysing, the possible interpretations and expectations concerning the analyses of hyperlinks in spatial research. The connections existing in cyberspace for selected self-government websites of Lower Silesia were shown to illustrate the issue. The analyses were conducted for selected websites of self-government units at all levels functioning in Lower Silesia. The paper presents two approaches to the analysis of hyperlinks: analysis of outlinks and inlinks. The presented results allow us to identify some regularities regarding the functioning of connections in cyberspace versus the connections in the real world. From the perspective of self-government websites it can be concluded that the connections in cyberspace reflect the real connections. In the majority of the analysed cases there is a clear connection with the actual scope of activity. The remaining relationships with administrative cities reflect the significance of these cities for the functioning of self-government units in the social, economic, legal, and administrative conditions. Based on the conducted analyses it can be concluded that the study of hyperlinks may be useful in understanding the relationships between geographical space and cyberspace. They may form a new, interesting field of spatial research. We also found some challenges in the study of hyperlinks' spatial aspects: the identification of consistent criteria for determining the websites' 'location in space', the possibility of interpreting the research results and the dynamics of the Internet.

Contents:

1. Introduction

The transfer of accents from geographical space to cyberspace in all spheres of life requires the search of new methods in the area of spatial research. It requires such methods which would help describe the new 'spatial' reference – the Internet. The elements which decide about the functionality of the Internet as a medium or a source of information are called hyperlinks. They are a kind of 'surfing board' which allows navigating the vast expanses of the Internet resources, and integrates them.

The purpose of this study is to show the selected ways of analysing, the possible interpretations and expectations concerning the analyses of hyperlinks in spatial research. Analyses of connections between websites that take into consideration their location are relatively rare in geographical research. Also, the studies conducted in other disciplines of knowledge relying more heavily on the analyses of hyperlinks (e.g. scientometrics) lack clearly defined spatial references. The spatial aspect is usually treated as peripheral in relation to the mainstream of the discussion.

The connections existing in cyberspace for selected self-government websites of Lower Silesia were shown to illustrate the issue. The seats of the entities responsible for the content of websites which were hyperlinked with websites of the Lower Silesian local and regional authorities were identified. The choice of self-government units was dictated by the possibility of determining the actual scope of activity in geographical space (administrative border) as well as the existence of more or less clearly formulated connections reaching beyond the scope of a single unit. It was assumed that with this choice it would be possible to compare the area of actual activity and the actual connections with the 'area' and the connections in cyberspace. This approach allows to test Tobler's First Law of Geography (TFL), formulated by Waldo Tobler: 'everything is related to everything else, but near things are more related than distant things' (Tobler, 1970: 236). Is it still valid in cyberspace?

2. Material and research methods

2.1. The essence of hyperlinks

The analysis of hyperlink connections between websites is the domain of webometrics. Webometrics can be defined as: 'The study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web drawing on bibliometric and informetric approaches' (Ingwersen, Björneborn, 2005: 341). Webometrics is also defined as a 'new kind of quantitative analysis' (Heimeriks, van den Besselaar, 2006: 2). In this case the main accent is on its character as a tool that can be applied in a variety of scientific areas.

Fig. 1 presents types of hyperlinks and the relationships that can occur between hyperlinked websites.
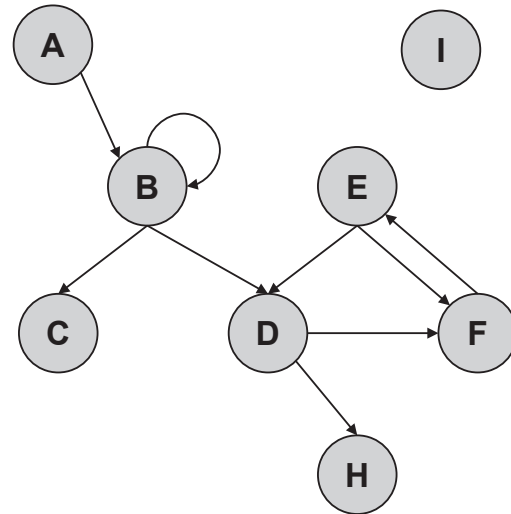
Fig. 1. Relationships between websites – types of hyperlinks (explanation within the text)

Source: Ingwersen, Björneborn, 2005

According to Ingwersen and Björneborn (2005) the basic terminology of webometrics is closely connected with the perception of the Internet through the prism of graphs, with reference to the analysis of social networks and bibliometrics. According to the authors mentioned above, the connections of the network nodes by hyperlinks can be described in the following way (with reference to Fig. 1):

- B has an inlink from A; B is inlinked; A is inlinking; A is an in-neighbour of B;
- B has an outlink to C; B is outlinking; C is an out-neighbour of B;
- B has a selflink; B is selflinking;
- A has no inlinks; A is non-linked;
- C has no outlinks; C is non-linking;
- I has neither in- nor outlinks; I is isolated;
- E and F have reciprocal links; E and F are reciprocally linked;
- D, E and F all have in- or outlinks connecting each other; they are triadically interlinked;
- A has a transversal outlink to G: functioning as a shortcut;

- H is reachable from A by a directed link path;
- C and D are co-linked by B; C and D have co-inlinks;
- B and E are co-linking to D; B and E have co-outlinks;
- Co-inlinks and co-outlinks are both cases of co-links.

Up until now the analyses of links were mainly used in scientometrics (e.g. Harries et al., 2004; Ortega, Aguillo, 2008a, 2008b, 2009; Ortega et al., 2008; Maggioni, Uberti, 2009; Park, 2010; Thelwall 2002a, 200b; Thelwall et al., 2003; Vaughan et al., 2007). In these cases the analysis of hyperlinks was used mainly to identify the connections between universities. However, it often did not take into account their location in geographical space. Such popularity of the analyses of hyperlinks in scientometrics results from an important characteristic of hyperlinks, i.e. they are equalled with bibliographical citations. The link that is placed on a particular website and directing to another one (outlink) should be identified as a reference to this website. Thus, the website that is linked is quoted (has an inlink). This type of cyberspace citation is often described as 'sitation' (Nowak, 2008). The differences existing between citations and sitations (hyperlinks) are mainly connected with the very nature of the Internet (see Scharnhorst, Thelwall, 2005).

Other fields of application of the analysis of hyperlinks are more aimed at the identification of spatial relationships. The works of Lin et al. (2007) and Takhteyev et al. (2011) take a different course – focusing on the connections between blogs. They identify the existence of significant dependencies between relationships in cyberspace and geographical space. The analyses of Halavais (2000) and Bharat et al. (2001) relate closely to the spatial aspects, and strive to identify the relationships between websites of entities from different countries. The analysis of hyperlinked sites and their thematic diversity is also useful in determining the connections between countries (Park, 2004). Hyperlinks may also be the basis for identifying the hierarchy of cities (Brunn, 2005). The analyses of hyperlinks are also used in research into the connections between local government websites (Holmberg, Thelwall, 2009; Holmberg, 2010).

2.2. Research procedure

The paper presents two approaches to the analysis of hyperlinks. First, the relationships were defined though the analysis of outgoing links (outlinks) from self-government websites of Lower Silesia. This approach allows us to identify the spatial distribution of destinations to which lead the links from a particular website. The other, complementary approach, is the analysis of incoming links to a particular website (inlinks). In this case, it takes into account the distribution of websites which link to a particular self-government website (the location of entities administrating them).

First of all, mirrors of all websites under research were obtained (program – HTTrack Website Copier). Then, one of the hyperlink extracting programmes was used to isolate hyperlinks from HTML code (outlinks) – Offline Link Extractor 1.7. The last stage was to visit each website connected with the analysed site by a link. Information about the geographical location of the headquarters of the institution, firm, host organization, or administrator of the site was obtained from the home page. In order to obtain information regarding inlinks advanced functions of the Google browser were used – the functions of finding websites which have links to particular website addresses. It needs to be emphasised that the Google browser is more and more often used in obtaining information for spatial research and concerns not only hyperlinks, but also information on the 'popularity' of some geographical phenomena (e.g. Brunn et al., 2010; Nunes, 2006). Obtaining information on location in geographical space was identical in the case of outlinks.

Official websites of local and regional authorities were chosen as the focus of the study. The analyses were conducted for selected websites of self-government units at all levels functioning in Lower Silesia. These included the websites of the Marshall's Office, Voivodship Office – the regional level; websites of poviats (administrative regions of the 2nd order) and gminas (administrative regions of the 3rd order) – at the local level. In the case of poviats, a special analysis was performed on two characteristic ones chosen on the grounds of the clearness of their relationships. The same was done in the case of gminas, but an additional criterion of selection was the existence of a relatively large number of links and the location within the poviats under analysis.

The study uses terminology referring to the theory of graphs. The websites of poviats and websites to which hyperlinks lead are regarded as nodes. The edges are hyperlink connections between the nodes, regardless of the number of hyperlinks. The analysis does not include nodes from outside Poland. In these cases the number of links (which constitute edges) for the websites under analysis did not exceed a few per cent.

3. Research results

First, hyperlink destinations for websites of institutions situated in Wrocław were identified. These included the Marshall's Office (UMWD), Voivodship Office (DUW) – the regional character – and the Municipality of Wrocław (Fig. 2). All the three websites are run by institutions which have different goals, scopes of activity and institutional connections. This is reflected in the network of nodes and edges and the number of links. The website of the Marshall's Office is strongly connected with the region. Out of

54 edges, 35 are connected to the nodes within Lower Silesia. As far as the number of links is concerned, Warsaw is undoubtedly in the lead (199), but the other most represented destinations are the main cities and towns of the region, i.e. Jelenia Góra (14), Legnica and Wałbrzych (9). The situation is different in the case of the Voivodship Office: the website is hermetic, with only 21 hyperlinks to Polish locations; lack of connection with the region – 1 edge. As far as the Municipality of Wrocław is concerned the connections lead mainly to Wrocław (682 links), with similar significance of destinations in the region (18 edges excluding Wrocław) and in Poland (20 edges). As for the number of links, apart from Wrocław,
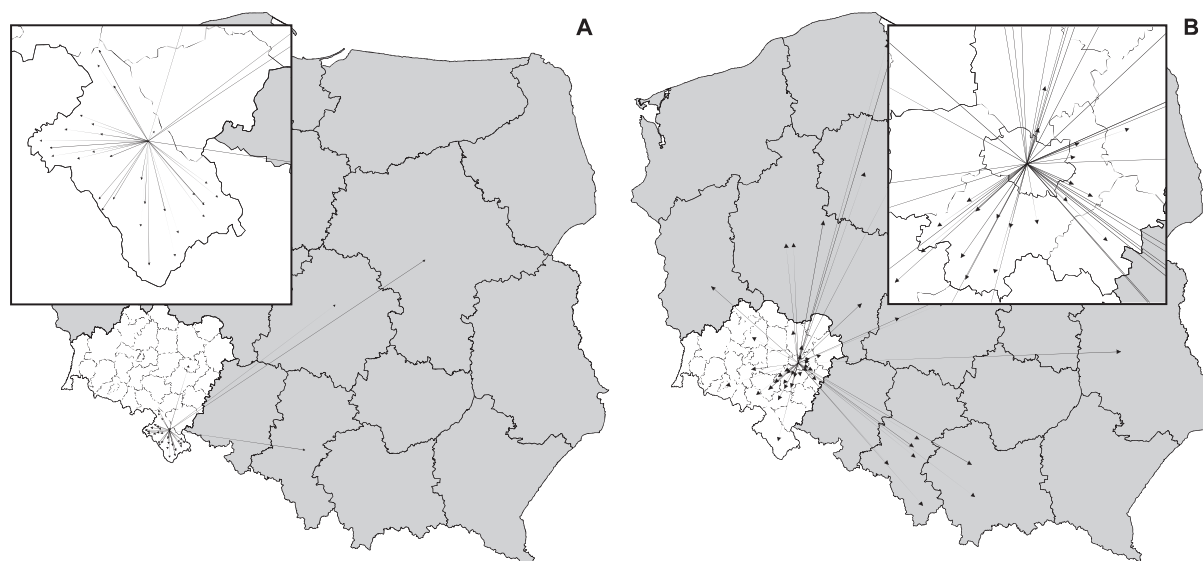

Fig. 2. Outgoing hyperlinks (outlinks) from websites

Explanation: A – Marshall's Office of Lower Silesia; B – Voivodship Office of Lower Silesia; C – Municipality of Wrocław

Source: Own study

Fig. 3. Outgoing hyperlinks from websites

Explanation: A – Kłodzko poviat; B – Wrocław poviat

Source: Own study

Warsaw is in the lead (105) together with other large cities from outside the region – Kraków (9), Poznań and Katowice (6).

An important criterion describing the character of the relationship is the number of links leading to Wrocław and the comparison of this number with the number of links leading to Warsaw. As was shown above, the Municipality of Wrocław is clearly orientated towards the capital of the region (69% of links) rather than the country capital (11%). The Marshall's Office is more strongly connected with Warsaw (30% of links compared to 15% directing to Wrocław). Even stronger orientation towards the country capital is observed in the case of the website of the Voivodship Office of Lower Silesia: 41% Warsaw – 11% Wrocław.

Another level of reference are poviats. From among 28 potential websites those belonging to the Poviat Authority of Wrocław and Kłodzko were chosen. The main feature distinguishing the two cases is the strength of the relationship with the area of the poviat. In the case of Kłodzko poviat the destination of links is strongly connected with the poviat area. Only 4 edges (among Polish locations) lead outside the region and 29 to locations within the poviat. The connections also indicate the significance of towns within the poviat. The greatest number of links connects the poviat website with Polanica-Zdrój (18), Duszniki-Zdrój (13) and Kudowa-Zdrój (12). These are spa towns which are important from the point of view of tourism as well as socio-economic potential.

The number of links is even greater than those to websites located in Kłodzko (8). The Wrocław poviat is characterised by two features. Firstly, a large concentration of links to Wrocław, and secondly, a tendency for the remaining connections to lead outside the region. In the case of edges the dominance is not so clear (a relatively good connection within rural gminas is observed), however, with respect to the number of links extra-region locations are most common: Warsaw – 121; Kraków – 9; Katowice – 6: Poznań – 6.

In the case of gminas, each of the selected examples represents different relationships. Kąty Wrocławskie (Wrocław poviat) is characterised by a very strong bond with the area of the gmina – out of 21 edges 10 are connected to the nodes of the gmina in question (mainly in the immediate surroundings of the Wrocław border). The towns within Wrocław poviat form 4 nodes. The website of Międzylesie gmina (Kłodzko poviat) is connected with 18 nodes in Poland. Only 4 of them come from Międzylesie gmina, and 4 from the other gminas of Kłodzko poviat.

From the perspective of links to nodes – administrative centres at various levels – Międzylesie is characterised by a balanced structure: 10% of the links lead to the capital of the gmina, 7% to the capital of the poviat, 10% to the capital of the region and finally 22% to the country capital. The links coming out of the town of Kąty Wrocławskie are mainly local links (43%), next comes Wrocław – the capital of the poviat (10%) and only 5% to Warsaw.
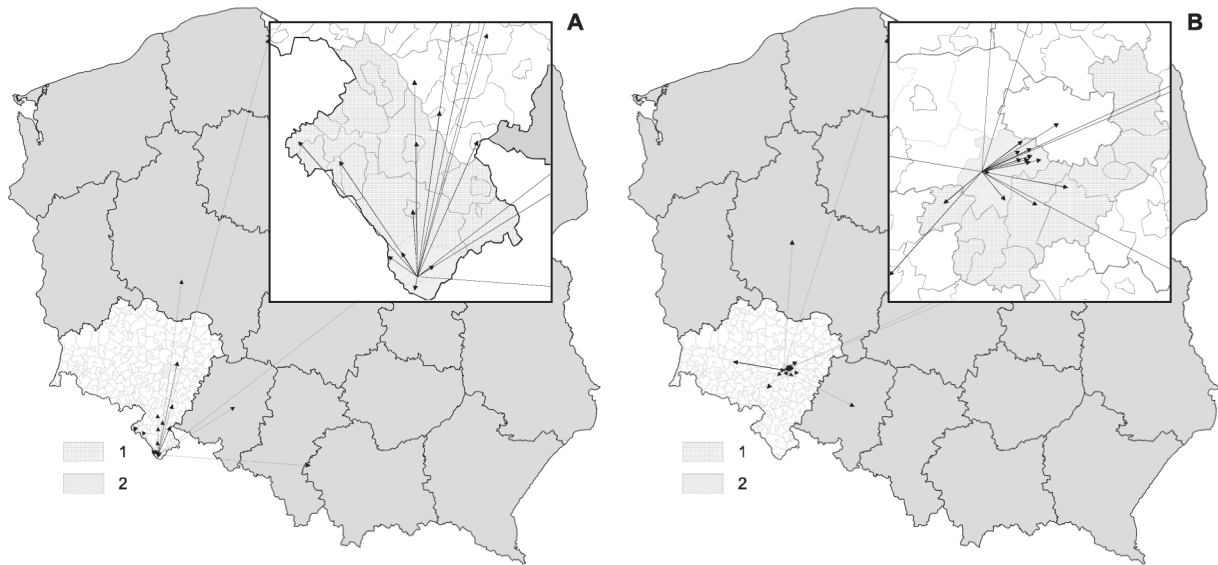
Fig. 4. Outgoing hyperlinks from the websites

Explanation: A – Międzylesie gmina; B – Kąty Wrocławskie gmina (marked 1 on the map – the area of the poviat; 2 – the area of a particular gmina)

Source: Own study

For the incoming links only those were presented which lead to the Municipality of Wrocław, and namely to its main website (www.wroclaw.pl). The connection with nodes from large cities from outside the region is noticeable, with a less pronounced participation of the nodes from within the region. Another characteristic feature is the dominance of links from websites 'situated' in Wrocław – 141 – with next nodes in Warsaw with 9 links, and Kraków with 7.

Fig. 5. Incoming hyperlinks (inlinks) to the website of the Municipality of Wrocław

Source: Own study

4. Discussion – hyperlinks in spatial research

The presented results allow us to identify some regularities regarding the functioning of connections in cyberspace versus the connections in the real world. From the perspective of self-government websites it can be concluded that the connections in cyberspace reflect the real connections. In the majority of the analysed cases there is a clear connection with the actual scope of activity (the borders of a given administrative unit). The remaining relationships with administrative cities reflect the significance of these cities for the functioning of self-government units in the social, economic, legal, and administrative conditions. In another study focusing on the hyperlink connections between the websites the author suggests that the relationships between all levels of self-government administration should be defined (see Janc, 2011). Having done this in the present study, it can be clearly stated that Warsaw is the central node of Polish cyberspace. The centrality of the country capital results from Warsaw being the administrative, political and cultural centre. You can also refer to the results of the study by Lin et al. (2007), which emphasises the existence of strong relationships between opinion-forming centres. Warsaw is the centre for the shaping of opinions, ways of thinking and acting. It is the seat of the most important non-governmental

organisations. Hence, its significant position as a node in cyberspace. The dominance of Warsaw is also noticeable in other aspects of the functioning of cyberspace – the flow of information between locations of Internet users (see Ilnicki, Janc, 2009). These conclusions correspond with TFL. Everything is still related, but we should notice that a strong relation does not require geographical proximity. It's fundamental from the perspective of new spatial logic – the space of flows (Castells, 1996). Another important finding is related to the idea of nodal regions. Nodal regions, as a special case of functional regions (e.g. Brown, Holmes, 1971), could be delimited by means of hyperlink analysis. The presented results show clear relations between geographical space in terms of functional relations (functional distance) and cyberspace (hyperlinks).

Taking into account the objectives set in this study, it is more important to interpret the results from the perspective of research opportunities hyperlinks offer geographical sciences rather than their substantial examination.

Firstly, with regard to the method of analysis, it needs to be noted that the study of hyperlinks, along with the entire webometrics, is a developing discipline. Apart from the well-known WIF (*Web Impact Factor*) there are no recognised methods of analyses – approaches, measures, indicators. From the point of view of spatial analyses at the early stage of development of this type of research it seems more significant to identify the 'spatial' relationships rather than developing new indicators. In this case, the achievements of webometrics should be used. In spatial analyses it seems sufficient to identify the phenomenon through the number of links and the distribution of nodes and edges. A reliable visualisation (map) will always be better in this case than the use of complex indicators. However, one of the major challenges is the identification of consistent criteria for determining the websites' 'location in space'. When conducting spatial research one needs to precisely define 'location' in cyberspace as well as in geographical space. Due to the fact that many websites are locationless (where do you situate Facebook, YouTube, Google?) or have multiple locations (does each profile on Facebook, Twitter form a different location? Does each language version of Google form a new location?) it seems extremely difficult to determine the locations and spatial relationships.

Another important issue is the possibility of interpreting the research results. The presented empirical analysis helped to determine the relationship between the actual and the virtual 'area'. Each of the presented means of connections between the websites (see Fig. 1) will have a different outcome. Their apparent asymmetry may create problems when analysing the results, especially in the case of studying connections for (between) websites of entities which do not have a clearly defined area of activity in geographical space (e.g. blogs, home pages of individuals and companies). The interpretation should also take into account the fact that the motives for creating hyperlinks are very different: e.g. they refer to an authority; support, approval of the contents and opinions; reinforcement of the message; trust; criticism (Ackland, 2010). Hence, the solution should be analysed outside the geographical location and thematic categories to which a hyperlink can be assigned (see Janc, 2011).

One should also mention the problems posed by 'spatial' analysis of hyperlinks. The analysis of inlinks is less accurate and does not fully reflect the reality. In this case, one should take into account the margin of error resulting from the functioning of the programmes indexing the Internet in search of hyperlinks. Each Internet browser yields different results. Another problem is the dynamics of the Internet. Changes in the resources, the appearance of some sites and the disappearance of others, the introduction of new versions of the same websites, the functioning of the so-called 'deep Internet' (not indexed by the web search engines), changes in the way search engines operate – all this may impede dynamic research. The situation cannot be compared at different times. Of course, in this case it is necessary to understand that we will probably never be able to determine the true size or structure of cyberspace. Any research conducted into cyberspace will only lead to approximating 'the reality', and not its faithful description. However, the same situation occurs in traditional geography – all the statistical data only help to roughly describe the actual processes and phenomena.

5. Conclusions – research directions

Based on the conducted analyses it can be concluded that the study of hyperlinks may be useful in understanding the relationships between geographical space and cyberspace. They may form a new, interesting field of spatial research. The examples presented in this paper are only a fragment of a larger whole, a fine element of the system of real connections. Categorical statements could only be made after examining the entire phenomenon (all self-government websites in Poland). However, due to the pioneering nature of

this research, only some generalisations were made which could be useful in the further development of this type of research. Therefore, we can identify the following potential fields of exploration of hyperlinks in spatial research: (a) relationships between different types of institutions aiming to identify interdependencies and connections – e.g. universities, cultural institutions, NGOs; (b) the analysis of the functioning of company websites. The possibility of defining connections within one company (subsidiaries, branch offices) and between enterprises (co-operation, interdependencies). One can also expect the possibility of identifying the potential/actual scope of the impact of a given company – in conjunction with the qualitative analysis of the websites (e.g. language versions, content analysis); (c) specifying the significance of 'places' and 'areas' in the hierarchy, the popularity of places and areas. The number of references (together with the in-depth analysis of the 'visibility' of the website on the Internet) to websites belonging to the entities located in a particular place may provide information about the role and interest in the place – something that is often impossible to determine through traditional statistical measurements assigned to the units of spatial reference.

References

Ackland, R. 2010: WWW hyperlink networks. In: Hansen, D., Shneiderman, B. and Smith, M.A. editors, *Analyzing social media networks with NodeXL: insights from a connected World*, Burlington: Morgan Kaufmann, pp. 181-199.

Bharat, K., Chang, B., Henzinger, M. and Ruhl, M. 2001: Who links to whom: mining linkage between Web sites. In: *Proceedings IEEE International Conference on Data Mining*, pp. 51-58.

Brown, L.A. and Holmes, J. 1971: The delimitation of functional regions, and hierarchies by functional distance approaches. In: *Journal of Regional Science*, 11, 1, pp. 57-72.

Brunn, S.D. 2005: An e-classification of the World's capital cities: URL references to Web sites. In: Kehal, H.S. and Singh, V.P. editors, *Digital economy: impacts, influences and challenges*, Hershey: Idea Group Publishing, pp. 200-216.

Brunn, S.D., Devriendt, L., Boulton, A., Derudder, A. and Witlox, F. 2010: Networks of European cities in worlds of global economic and environmental change. In: *Fennia*, 188, 1, pp. 37-49.

Castells, M. 1996: The rise of the network society, Oxford: Blackwell Publishers Ltd.

Halavais, A. 2000: National borders on the World Wide Web. In: *New Media Society*, 2, 1, pp. 7-28.

Harries, G., Wilkinson, D., Price, L., Fairclough, R. and Thelwall, M. 2004: Hyperlinks as a data source for science mapping. In: *Journal of Information Science*, 30, 5, pp. 436-447. DOI: 10.1177/0165551504046736.

Heimeriks, G. and van den Besselaar, P. 2006: Analyzing hyperlinks networks: the meaning of hyperlink based indicators of knowledge. In: *International Journal of Scientometrics, Informetrics and Bibliometrics*, 10, 1, pp. 1-17.

Holmberg, K. 2010: Co-inlinking to a municipal Web space: a webometric and content analysis. In: *Scientometrics*, 83, pp. 851-862. DOI: 10.1007/s11192-009-0148-1.

Holmberg, K. and Thelwall, M. 2009: Local government web sites in Finland: A geographic and webometric analysis. In: *Scientometrics*, 79, 1, pp. 157-169. DOI: 10.1007/s11192-009-0410-6.

Ilnicki, D. and Janc, K. 2009: Topology, nodality and space of internet flows. In: *Acta Universitatis Palackianae Olomucensis – Geographica*, 40, 2, pp. 15-26.

Ingwersen, P. and Björneborn, L. 2005: Methodological issues of webometric studies. In: *Handbook of Quantitative Science and Technology Research, part 2*, Springer, pp. 339-369.

Janc, K. 2011: Geografia hiperlinków – przestrzenny wymiar samorządowych serwisów internetowych. In: *Studia Regionalne i Lokalne*, 4 (46), Warszawa: Wydawnictwo Naukowe SCHOLAR, pp. 30–50.

Lin, J., Halavais, A. and Zhang, B. 2007: The blog network in America: blogs as indicators of relationships among US cities. In: *Connections*, 27, 2, pp. 15-23.

Maggioni, M.A. and Uberti, T.E. 2009: Knowledge networks across Europe: which distance matters? In: *Annals of Regional Science*, 43, pp. 691-720. DOI: 10.1007/s00168-008-0254-7.

Nowak, P. 2008: Bibliometria. Webometria. Podstawy. Wybrane zastosowania., Poznań: Wydawnictwo Naukowe UAM.

Nunes, F. 2006: The Portuguese urban system: an opposition between its hierarchical organization in cyberspace vs. physical space. In: *Telematics and Informatics*, 23, 2, pp. 74-94. DOI: 10.1016/j.tele.2005.05.001.

Ortega, J.L. and Aguillo, I.F. 2008a: Linking patterns in European Union countries: geographical maps of the European academic web space. In: *Journal of Information Science*, 35, 5, pp. 705-714. DOI: 10.1177/0165551507086990.

Ortega, J.L. and Aguillo, I.F. 2008b: Visualization of the Nordic academic web: link analysis using social network

tools. In: *Information Processing & Management*, 44, pp. 1624-1633. DOI: 10.1016/j.ipm.2007.09.010.

Ortega, J.L. and Aguillo, I.F. 2009: Mapping word-class universities on the web. In: *Information Processing & Management*, 45, pp. 272-279. DOI: 10.1016/j.ipm.2008.10.001.

Park, H.W. 2004: Presence of Taiwan on the World Wide Web in South Korea: dynamics of digital and geographical presence on cyberspace. In: *The International Information & Library Review*, 36, 4, pp. 329-340. DOI: 10.1016/j.iilr.2003.12.001.

Park, H.W. 2010: Mapping the e-science landscape in South Korea using the webometrics method. In: *Journal of Computer-Mediated Communication*, 15, pp. 211-229. DOI: 10.1111/j.1083-6101.2010.01517.x.

Scharnhorst, A. and Thelwall, M. 2005: Citation and hyperlink networks. In: *Current Science*, 89, 9, pp. 1518-1524.

Takhteyev, Y., Gruzd, A. and Wellman, B. 2011: Geography of Twitter Networks. In: *Social Networks*, 34, 1, pp. 73-81. DOI 10.1016/j.socnet.2011.05.006.

Thelwall, M. 2002a: The top 100 linked-to pages on UK university web sites: high inlink counts are not usually associated with quality scholarly content. In: *Journal of Information Science*, 28, 6, pp. 483-491. DOI: 10.1177/016555150202800604.

Thelwall, M. 2002b: Evidence for the existence of geographic trends in university Web site interlinking. In: *Journal of Documentation*, 58, 5, pp. 563-574. DOI: 10.1108/00220410210441586.

Thelwall, M., Tang, R. and Price, L. 2003: Linguistic patterns of academic Web use in Western Europe. In: *Scientometrics*, 56, 3, pp. 417-432.

Tobler, W.L. 1970: A computer movie simulating urban growth in the Detroit Region. In: *Economic Geography*, 46, Supplement, pp. 234-240.

Vaughan, L., Kipp, M.E.I. and Gao, Y. 2007: Why are websites co-linked? The case of Canadian universities. In: *Scientometrics*, 72, 1, pp. 81-92. DOI: 10.1007/s11192-007-1707-y.