

Social networks in spatial data infrastructures

P. A. J. van Oort · G. W. Hazeu · H. Kramer ·
A. K. Bregt · F. I. Rip

Published online: 12 June 2009

© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract Data sharing over the internet is increasingly efficient and increasingly anonymous. This growing anonymity makes it difficult for data users to exchange feedback: interesting ideas, innovations, ancillary data, concerns about data quality, etcetera. For data producers not knowing the end-users makes it more difficult to collect end-user feedback and tailor the production process to users needs. Feedback can be exchanged through social networks, in other words contacts with other users. Very little is known about such social networks in the case of data sharing. In this paper we present an analysis of the social networks

associated with two land cover datasets. We found that most users have zero to two links to other users and virtually no links to users in other organisations. Without these links, it seems almost impossible to identify users of a particular dataset. Internet technology can help but is currently not used to enhance exchange of feedback within the user community.

Keywords Social networks · Spatial data infrastructure · User · Producer

Introduction

The internet revolution is changing the way we exchange information. Instead of going to a person to get access to a particular dataset we now simply download it.¹ At the expense of increased efficiency we also see increasing anonymity. It becomes increasingly difficult to identify end-users (Morville 2005). The internet revolution has also increased the efficiency with which people can manage their social ties, through community websites and email. However, the existence of such social ties in response to growing anonymity is not obvious. If the effort for maintaining these ties is considered too large and/or

P. A. J. van Oort (✉) · G. W. Hazeu · H. Kramer ·
A. K. Bregt · F. I. Rip
Centre for Geo-Information, Wageningen University,
P. O. Box 47, 6700 AA Wageningen, The Netherlands
e-mail: pepijn.vanoort@wur.nl

G. W. Hazeu
e-mail: gerard.hazeu@wur.nl

H. Kramer
e-mail: henk.kramer@wur.nl

A. K. Bregt
e-mail: arnold.bregt@wur.nl

F. I. Rip
e-mail: frans.rip@wur.nl

Present Address:

P. A. J. van Oort
Centre for Cropping Systems Analysis, Wageningen
University, P.O. Box 430, 6700 AK Wageningen,
The Netherlands

¹ This applies to data freely distributed over the internet as well as datasets freely accessible from an internal server within an individual organisation.

the benefits of it are considered to be too low it may not happen. Especially in data infrastructures we expect a growing anonymity, because:

1. The social ties (professional, dataset specific, ad-hoc) are weaker than social ties like friendship or kinship;
2. People and organisations may not yet have become aware of the need of more explicitly managing these ties in the face of growing anonymity;
3. Users are often scattered over many organisations or different units within a larger organisation; consequently the chances that users of the same dataset meet each other to exchange ideas are limited;
4. For obvious reasons of privacy and bureaucracy personal details are often not registered at the dataset level, in which case it is impossible to identify users by following the data flow.

Such a set-up in which end-users are largely unknown potentially has some serious disadvantages:

1. Quality control: Producers cannot improve the quality of their product by asking their end-users for feedback if they don't know their end users;
2. Justification of funding: If we don't know who is using a dataset then we also don't know how many people are actually using it. Justification of funding towards budget managers becomes problematic in this light;
3. Data management: Data managers are responsible for a complete coverage of relevant datasets for their organisation. Not knowing the end-users (and their information needs) also makes it difficult to anticipating their needs in terms of securing data availability;
4. Knowledge transfer: users cannot share amongst each other interesting ideas, innovations, ancillary data,² and concerns with regard to a specific dataset if they lack links to other users.

Are the consequences of growing anonymity as listed above really a problem? That depends on two factors: (1) the magnitude of these consequences relative to other problems and (2) the degree to which users of a particular dataset exchange feedback through

their social network (of users the same dataset). For example if there are no concerns about data quality and no perceived benefit in knowledge transfer then users may feel less the need for exchanging feedback. For the potential users who lack access, data-sharing is their key problem. Improved access albeit with growing anonymity may be very well acceptable to such users. Indeed in the past decades spatial data infrastructure (SDI) research and policy making has been very much focussed on (enhancing) the flow of spatial data from producer to user (Clinton 1994; Azad and Wiggins 1995; Campbell and Masser 1995; Nedovic-Budic and Pinto 1999; Nedovic-Budic and Pinto 2000; Groot and McLaughlin 2000; Wehn de Montalvo 2003; Williamson et al. 2003; Masser 2005; Harvey and Tulloch 2006; De Vos 2007; Goodchild et al. 2007; Omran and van Etten 2007; Elwood 2008; www.ec-gis.org/inspire). The novelty in voluntary geographic information (VGI) is that it recognises that not only traditional mapping agencies but also individual citizens can have the role of spatial data producer (Budhathoki et al. 2008; Goodchild 2008). But also the emerging research on VGI is on spatial data (exchange, quality, etcetera) and not on the related issue of sharing of feedback. Sharing of spatial data is often impeded by cultural, financial, organisational and technical barriers. Most research and policy making over the past two decades has gone into understanding and then breaking down these barriers.

With progress being made in spatial data sharing attention is shifting towards how users exchange feedback. This aspect of data sharing has to date not at all been systematically studied. We identify three directions of feedback: user-to-user, user-to-producer and producer-to-user. Exchange of feedback has traditionally been through direct contact between users/producers, in meetings or via email. Exchanging feedback can be mediated through social networking media such as facebook, flickr, twitter, hyves or LinkedIn, discussion fora on a product-website (Scharl and Tochtermann 2007). To our best knowledge such social networking media have to date not been taken up by users for exchanging dataset-specific feedback, nor are such media supported or introduced by producers of spatial data. We therefore focus on direct contact between users/producers. If users cannot name other users (through their social network) then this implies very limited opportunities for exchanging feedback. The key questions addressed in this research are:

² For example for users of satellite images covering a specific part of the world data previously collected by other users in the same area can be very useful for calibration. And more generally in identifying fellow researchers active in the same area.

1. what do the social networks of users of a specific dataset look like? and
2. in what way is feedback exchanged through these networks?

The questions are answered for two particular datasets. For these two datasets we explore through surveys the user population. Survey non-response is also studied and its consequences for research outcomes discussed.

Materials and methods

Survey procedure and datasets

Survey procedure

Starting the survey we had an incomplete list of users and no insight in links between users. The snowball procedure (Hanneman and Riddle 2005) is in this case the appropriate method to obtain a map of the social network. One starts with an incomplete list of users. These users are requested to name new users (making use of their specific links in the network which are at that stage unknown to us researchers). This iterative procedure stops when no more new names are mentioned. The result is ideally a complete map of all links between all users. We started by mining user registrations provided by the producers and their salesperson, using registrations from January 2004 to August 2007. Survey questions Q3, Q7 and Q8 (appendix) were used as name generators (Marsden 2005). In case of non-response a first reminder was sent after 1 month and if necessary a second one after 2 months. Once the snowball came to a halt, we still did not know whether it yielded us information on the complete user population. We do not know if someone who did not respond is or is not a user. We do not know if all users were mentioned. Non response from a single contact in a particular organisation would leave the complete user population in that organisation undetected. We present a discussion on the non-response in section “[Survey non-response](#)”.

Social network analysis

We are aware of a whole body of literature on quantitative methods in social network analysis (e.g.

Carrington et al. 2005; Hanneman and Riddle 2005). Our choice was to use them only very sparingly because the networks that we discovered were small and clear enough for visual interpretation.

Geo-datasets

The two datasets analysed are actually two product families. LGN is the national land cover database of the Netherlands. LGN is a product family. The first dataset (LGN1) was delivered in 1986 and updates have since appeared every 3–5 years. LGN is a 25 m resolution raster dataset with 5 main and 39 sub land cover classes. LGN is used in a wide range of environmental applications. User conferences have been organised to receive feedback from users. There is extensive documentation (Thunnissen et al. 1992 a, b; Hazeu 2005) and several scientific publications (de Bruin et al. 2004; de Wit and Clevers 2004; van Oort et al. 2004). Metadata is delivered to users when buying the dataset and is available from www.lgn.nl.

HGN is the historical land cover database of the Netherlands. It is a product family, comprising the following datasets: HGN1900, HGN1960 and HGN1990. They are scanned historical topographical maps (georeferenced, 1850–1935) and more recent topographical maps (1:25.000, 1940–1994). The first HGN products were delivered in 2004. In comparison with LGN, the user community is smaller. There have been no user surveys or conferences thus far. People or organisations buying HGN receive a report describing in detail the production method. For further details we refer to Knol 2003, 2004) and www.hgnnederland.nl.

Charged data

LGN and HGN are charged datasets, i.e. individual users or organisations pay for access. One implication of this is was that for both datasets the salesperson had kept a record of buyers, which was convenient in getting the snowball procedure started. For spatial data distributed free of charge over the internet with no form of registration collecting an initial list of users can be more difficult though not impossible.

Charged data implies a risk of non-response from illegal users. We included in the introduction letter to the survey a subsection raising this issue and giving respondents four instructions on this issue: (1) If you are uncertain about whether or not you are an illegal

user you might first check internally with your legal officer; (2) all major users (such as provinces & ministries and most waterboards) all have a license at organization level so all their employees are legal users; (3) in some cases private companies get a project license: use of LGN beyond the duration of the project is illegal (LGN is to be deleted when the project ends). If this applies to you and if you had at any point in time such a project license then you can safely fill in; (4) of course you can always choose not to fill in your contact details. With these instructions and under these conditions (major users are all legal users) illegal use was not considered a major issue. It is, however, a relevant point of consideration when replicating this research for other changed data sets. And in surveys like this being open and clear on this issue towards respondents is essential.

User definition

The definition of the term “user” determines the extent of the user population. Many decision makers use reports based on data without being aware of the underlying data (Masser et al. 2008). We refer to this group as unaware users. Unaware users were excluded from the survey. We classified our respondents according to how frequently they produce images/graphs/tables for unaware users (appendix, survey question Q4). We identified 4 user roles (Q2) and acknowledged that one person can have multiple roles:

1. Intermediary = data manager, buyer/salesperson;
2. Direct user = I work directly with the geo-dataset;
3. Indirect user = I do not work directly with the geo-dataset. But I do know the dataset and it is used in processes in which I am involved;
4. Ex-user = I did at one time use the dataset but not during the past year.

Results

Response

From September 2007 to February 2008 we compiled a list of 339 email addresses of people named as LGN user and 91 names of people who might be HGN user. Table 1 shows their status. We use the term “real” user for all people who did respond and who are not ex-users. According to this definition we have

Table 1 Response status

	LGN	HGN
Open (=no response after three requests)	156	33
Closed, filled in survey		
With ex-users	108	39
Without ex-users (=real users)	94	36
Does not consider him/herself a user	49	11
Email request to fill in survey bounced	17	6
Does not want to respond	9	
Bought the dataset, but not yet using it		2
Total	339	91

94 (LGN) and 36 (HGN) real users. Non-response is discussed in section “[Survey non-response](#)”.

Results indicate that both datasets are being used in many different organisations (Table 2) and that many people in our list did not respond (LGN response rate: 41%, HGN: 54% see footnote.³). Non response at the organisation level is presented in Table 2: our full list of organisations that possibly host LGN users contained 54 organisations. We got a response from 32 (59%) of these organisations. Both datasets have one single large organisation that hosts ca 45% of all users: Wageningen University and Research Centre. According to Table 2 around 80% of the organisations have 3 or less users, around 65% have only one user.

Users

Table 3 describes the user population. LGN has a relatively high percentage indirect users and HGN has the highest percentage direct users. LGN users report relatively more frequently to unaware users. It suggests that LGN is more strongly embedded in decision-making processes and/or models.⁴ This could be explained from the age of the datasets—LGN exists longer so it has had more time to become embedded. Also a more specialised product like HGN (historical land cover) may require more time to find its way into applications.

³ Calculated from Table 1 as $108/(108 + 156) = 41\%$ for LGN and $39/(39 + 33) = 54\%$ for HGN.

⁴ This is also supported by the fraction of users who do no GIS operations on the dataset before working with it (Q10 — appendix): 23% for LGN and 41% for HGN.

Table 2 LGN/HGN using organisations

	LGN	HGN
Total mentioned (including non response)	54	18
Total response (organisations from which at least 1 person responded)	32 (59%)	12 (67%)
Responding organisations with 3 or less real ^a users	27 (84%)	9 (75%)
Responding organisations with only 1 real ^a user	20 (63%)	8 (67%)
Major LGN/HGN using organisations	% of population	
Wageningen UR	44	46
Water boards (14 in total)	25	
National forestry agency		13
Netherlands environmental assessment agency	8	15

^a Recall we defined “real” users as those who did respond and who are not an ex-user

Table 3 Users

	User roles				Reporting to unaware users			
	Intermediary	Direct	Indirect	Sum	Frequently	Sometimes	Never	Sum
LGN	29 24%	61 50%	31 26%	121 100%	12 11%	55 51%	40 37%	107 100%
HGN	11 28%	24 60%	5 13%	40 100%	1 3%	17 45%	20 53%	38 100%

Cross-tabulation of the two classifications in Table 3 revealed that the two are uncorrelated. We also checked if the number of outgoing links differed between respondents according to their user role. When considering all names mentioned (including non-response) one finds both for LGN and HGN that intermediaries, in comparison with other user roles, have more links to other users (Tables 5, 6). However, this difference disappears when accounting for non-response which is also higher for the intermediaries (§3.5.1). If we count only the links to real users (those people who did respond and who are not ex-users) we find no difference between users roles in terms of number of links to other users.

Considering the high number of organisations with less than three LGN/HGN users (Table 2) a comparison between organisations in terms of user roles was impossible. For LGN the population was large enough to categorise organisations (Table 4). The organisation Wageningen UR as a single user has a fraction of intermediaries of 0.13 as opposed to around 0.33 for most organisation categories. It suggests that this organisation has an effective SDI in place. Possibly this is related to the size of the organisation and the number of GIS users within the

Table 4 User roles per organisation category for LGN

	<i>N</i>	Intermediaries	Direct users	Indirect users
Wageningen UR (WUR)	50	0.13	0.53	0.34
Education/research (excl. WUR)	15	0.29	0.57	0.14
Waterboards	27	0.33	0.55	0.12
Provinces	10	0.36	0.27	0.36
Ministries	4	0.33	0.50	0.17
Private companies	4	0.25	0.25	0.50
Municipality	1	0.50	0.50	0.00

organisation. Water boards have relatively many direct users and relatively few indirect users. This is in line with the more operational role of these organisations in Dutch land use planning and water management. Provinces have a more coordinating role, as an intermediary between national policies and implementation at municipal/waterboard level. The relatively high fractions of intermediaries and indirect users are in line with this role. For the other categories, the number of respondents seems too low to draw conclusions.

User interaction

We studied user interaction in terms of (1) users giving feedback to the producer (2) whether people needed personal help from others in accessing data and (3) the prime source of metadata: using available metadata versus asking another user. Of the respondents 34%/42% (LGN/HGN) has at one point in time given feedback to the producer. If these respondents give feedback also on behalf of all other users in their organisation then this covers for LGN ca 75% of the user population. Cross-tabulation with user roles revealed that around 50% of the intermediaries and direct users gave feedback, as opposed to 18% of the indirect users. Overall, the results suggest that it is mostly the organisations with a small number (1 or 2) of LGN/HGN users and the indirect users who are not giving feedback to the producer. With regard to help with access (2) we found for both datasets that 52% of the users needed no help with accessing metadata. With regard to metadata (3) we found that 79%/89% (LGN/HGN) used written metadata (as opposed to human help) as the prime source of metadata. We think these numbers are high in comparison with what may be found for other datasets, for a number of reasons: the nature of the user population (professional), the state of information systems within their organisation (advanced) and the amount of available written metadata on these two particular datasets which is extremely high. In the literature, the lack of written metadata as well as organised user communities for sharing and building up a metadata knowledge base has been noted as a serious problem (Engler and Hall 2007).

Network analysis

Figures 1 and 2 show the social networks of LGN and HGN based on answers to survey questions Q3, Q7 and Q8. Colours are for different organisations, squares are users with the intermediary role and circles are users without the intermediary role. Figures 1, 2, and 3 show only the real users: users who did respond and excluding pure ex-users. Tables 5 and 6 are based on counts of all names generated, including those of people who did not respond. Similarities in the two networks (Figs. 1, 2):

- Both networks have two central nodes: the producer (left) and the salesperson.⁵ (right). Together these two have links to 65%/58% (LGN/HGN) of the populations depicted in Figs. 1 and 2. We expect these numbers are exceptionally high when compared with datasets that are distributed free of charge over the internet.⁶
- The number of links per user follows a power law distribution (Fig. 3) with few nodes having a high number of links and the majority a very low number of links. Of the respondents 60%/40% mention 0 other users, 85% mentions 2 or less other real users. Albert et al. (2000) have shown such networks are vulnerable to fragmentation when central nodes are removed. In our case a decision by the producer/salesperson to no longer register users corresponds with a central node removal and would severely fragment the network. Such a decision would make sense in terms of reducing bureaucracy but would at the same time severely limit users' opportunities to find and interact with users;
- Many (46%) of the users have no links to other real users within their own organisation;
- Most users have no external links (LGN 81%, HGN 50%);
- In the majority of cases the only external link is between the LGN/HGN using organisation and the producer. There are virtually no links between LGN/HGN users in different organisations;
- Contrary to expectations, the number of links from intermediaries to real users is not higher than the number of outgoing links of non-intermediaries. This surprising result is further addressed in section "Accuracy";

⁵ For clarity, the salesperson is depicted as one node. In reality it is a help desk (GeoDesk) run by a team of 5 people working part time at this helpdesk. Geodesk is responsible for internal SDI management, distribution of GIS and Remote Sensing software, models, geodata and metadata.

⁶ The high percentages are largely due to the fact that both datasets are commercial datasets. Intermediaries at other government agencies buy a license and then freely place LGN/HGN on their internal server. Geodesk therefore keeps an administration of external licenses. Up to 2007 Geodesk also individually registered LGN/HGN users within Wageningen UR.

Fig. 1 Sociogram of the LGN social network, only the real users *Colours* represent different organisations. Users with the intermediary role are indicated as *squares*, others as *circles*. Except for the *blue colour* (Wageningen UR), *colours* and *numbers* do not correspond with those in Fig. 2. Figure drawn with Netdraw (Borgatti 2002). (Color figure online)

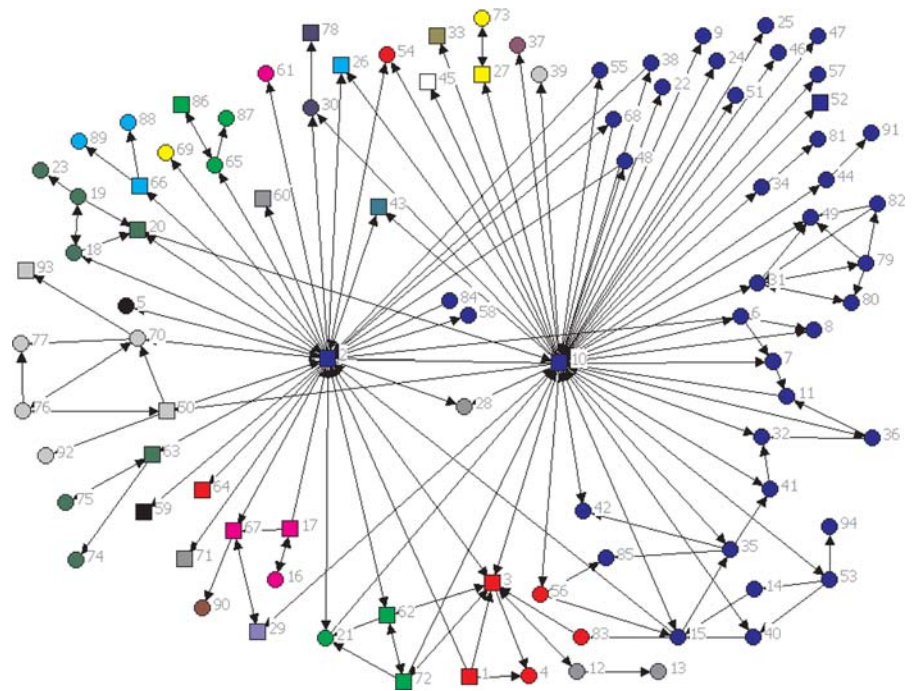
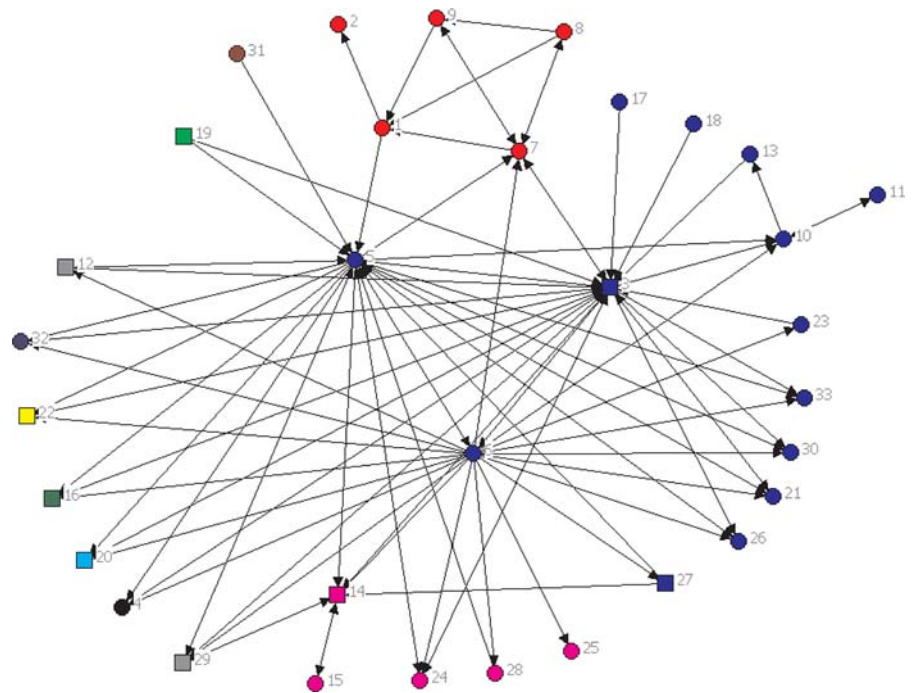


Fig. 2 Sociogram of the HGN social network, only the real users *Colours* represent different organisations. Users with the intermediary role are indicated as *squares*, others as *circles*. Except for the *blue colour* (Wageningen UR), *colours* and *numbers* do not correspond with those in Fig. 1. Figure drawn with Netdraw (Borgatti 2002). (Color figure online)



- It may seem strange to find so few linkages internally between users inside a single organisation like Wageningen UR (WUR). However, it is not so strange if we consider that WUR is a very large organisation with a very broad scope of

research. The part of the organisation in which people use maps had ca 926 employees at the time of this research. The 94 real LGN users (Table 1) represent only 10% of the organisation, for HGN the percentage is even lower. These people need

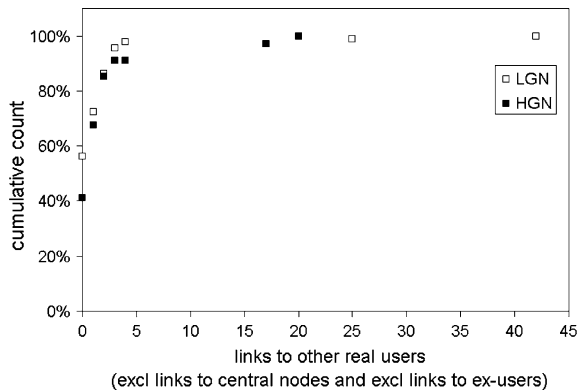


Fig. 3 Number of links to other real* users (=outgoing links) * recall we defined “real” users as those who did respond and who are not an ex-user

not work together, they may be active in different scientific disciplines in different sub-units of the organisation.

Notable differences between the two networks:

- HGN has one extra central node. This is a person who used the HGN producers’ list of users to advertise a report on a topic related to HGN. He has the same number of outgoing links as the other two central nodes, but almost no incoming links (as can be seen from the arrowheads). That

is: this person is not mentioned as a user by the other users;

- Within Wageningen UR there are almost no links between HGN users (only to and from the producer/salesperson). For LGN within Wageningen there are more links between users.

Survey non-response

For the large fraction of people who did not respond (Table 1) we simply do not know whether they are or are not real users. Accuracy, recall and non-response are important methodological concerns in social network analysis (Brewer 2000; Marsden 2005). We will look into two extreme possibilities:

- Accuracy: assume that all non-respondents are false recalls, i.e. the non-respondents are not real users. This means that our respondents often replied inaccurately to our request for names;
- Completeness: assume that all persons in our database are real users. In that case non-response results in an incomplete map of the social network. If a real user is not recalled by any of the users in our network we will remain unaware of the existence of this user. Especially where there are few social ties (as in Figs. 1, 2), poor

Table 5 Names mentioned by LGN users

User role(s) ^a	N	Same organisation		Other organisation		Total names mentioned		Accuracy μ (%) ^c
		>0 (%)	μ^b	>0 (%)	μ	>0 (%)	μ	
1	9	78	4	44	0	89	5	19
2	42	67	2	19	0	71	2	50
3	20	85	2	5	0	85	2	50
12	7	43	0	57	1	57	2	58
13	2	100	3	50	2	100	4.5	29
14	2	100	3	50	2	100	4.5	25
23	3	33	0	33	0	67	3	46
123	3	33	0	67	3	67	3	54
234	1	100	1	0	0	100	1	0

^a 1 = intermediary, 2 = direct user, 3 = indirect user, 4 = ex-user; note that users may have more than one role. Example: user role 12 is people with both the intermediary (1) and the direct user (2) role

^b >0 (%) is the percentage of respondents that returned at least 1 name of another user; μ is the median number of names mentioned. Thus out of $N = 9$ intermediaries (user role 1) 89% mentioned 1 or more names of other users. Most intermediaries generated 5 names

^c μ (%) is the median accuracy in percentage. For a respondent who mentioned 6 user names of which 2 responded we calculated the accuracy as $2/6 = 33\%$. μ (%) is the median of all individual respondents’ accuracies. Accuracy only calculated for respondents who generated one or more names

recall can contribute to an incomplete mapping of the network.

We explore these two extremes in the following subsections.

Accuracy

Assuming that those who did not respond are no real users we calculated for each respondent his/her accuracy. Imagine respondent A who mentioned 4 names. To these 4 we also sent our survey and only 1 responded. In that case the accuracy is set to $1/4 = 25\%$. The rightmost column in Tables 5 and 6 shows the median of all individual respondents' accuracies. Table 5 shows that intermediaries mention more names than other users (LGN median (μ) 5 vs. 2), but their accuracy is also lower (LGN median $\mu(\%)$ 19% vs. 50%). Consequentially, the number of links to real users is not much different between the user roles. Similarly for HGN (Table 6) we find that intermediaries generate more names and have a lower accuracy.

We hypothesise that the low accuracies are at least partially attributable to the level of recall: organisation (or unit within) versus individual (cf. Hansen 1999). While the people who actually use LGN/HGN may change (change of job or tasks) the organisation (unit) may be a constant user. In that case it makes more sense for users to recall the organisation (or unit) and less sense to recall the individuals. This hypothesis is supported by closer analysis of the results of our request for names (Table 7): despite the explicit request for personal details, many respondents returned names of organisations (or units within). Another eligible unit of analysis could be

models that use LGN/HGN as input. Discussion of our survey results with the user community suggested that people discuss with each other and recall what models they are working on, without exchanging specifics on input data. Just like organisational units, models may act as an entry point to identifying user communities that we shall pursue in further research.

Completeness

The other extreme possibility is that all those who did not respond are in fact real users. Under this assumption, we have captured only 59%/67% of the organisations using LGN/HGN (Table 2) and only 41%/54% of the people using LGN/HGN. Results, however, may be even worse. A detailed list of the response was reviewed by the data producers and they pointed to certain organisations that were missing from the list and organisations for which a higher number of users was expected. The general methodological problems in identifying users as noted in the introduction of this paper remain: without any lead into an organisation, or with just one lead who cannot or will not cooperate, finding more users is like looking for a needle in a haystack. For the network topology (Figs. 1, 2, 3) that we found, with very few linkages to users, we are very much dependent on individual response. Networks with a more homogeneous topology (Albert et al. 2000) are less sensitive to non-response.

Figures 1 and 2 suggest that if users would like to contact other users, then finding them is almost impossible without help from the central nodes. In fact, such an opportunity arose shortly before the survey started. In June 2007, an LGN user conference was organised. One would expect that if users saw a

Table 6 Names mentioned by HGN users

User role(s)	N	Same organisation		Other organisation		Total names mentioned		Accuracy μ (%)
		>0 (%)	μ	>0 (%)	μ	>0 (%)	μ	
1	7	57	1	86	2	86	3	46
2	19	74	1	21	0	79	1	100
3	2	0	0	50	1	50	0.5	100
12	3	67	3	100	1	100	5	40
23	2	50	14	100	11	100	24	41
34	1	0	0	100	1	100	1	100

Meaning of column headings is the same as in Table 5

Table 7 Request for names of other users

	LGN	HGN	Total
One or more names of other users	59	20	79
Names of organisations (or units within), no names of individuals	21	6	27
“Students”	7	1	8
On an internal server, not a clue who is accessing it	4	1	6
Don’t want to give names	2	1	3
Not yet using the geo-dataset	0	2	2
Other namely	1	1	2

need or benefit in user interaction, then this would be their chance. However, in our survey results, we found that apart from one exception conference attendees gave no names or only a fraction of the names of other conference attendees. The single exception was the LGN producer who organised the conference and did clearly recognise the importance of feedback from the users. It is unclear at this stage why other users did not recall names of other conference attendees when filling in our survey. Possibly users rely on the producer in continuing this role of mediator so that no additional effort is required from the users.

Discussion

We studied three categories of linkages between users: (1) producer to user (2) user to producer and (3) user to user. The energy of SDI programs (Clinton 1994; Groot and McLaughlin 2000; Williamson et al. 2003; Masser 2005; www.ec-gis.org/inspire) seems to be going mainly to increasing the efficiency and effectiveness of (1) with little interest in the role of users in SDIs (Budhathoki et al. 2008). SDI technologies, standards and policies are improving the efficiency of the transfer of data/metadata, so that users find it increasingly easy to find and access data and metadata. Progress in these fields in the Dutch spatial data infrastructure (VROM 2008) is reflected in our results where a large fraction of users required no help in accessing data (52%) or metadata (around 80%). The flow from user to producer is also present: around 35% of the users did at one point in time give feedback to the data producer; if on behalf of other staff in their organisation this covers 75% of the user

population. This is probably quite high in comparison with datasets that are exchanged freely over the internet.

What is really absent is the user to user interaction. For two datasets we found that, discarding links to the producer, around 50% cannot name a single other user; the majority of users (88%) can only directly name 2 or less other real users. There were virtually no links between users in different organisations. Users often have an idea of which organisations or sub departments are likely users, but lack at a dataset level direct links to other users. For geo-datasets distributed freely over the world wide web we anticipate it to be even worse, with users having virtually no ties to other users of the same dataset, nor opportunities to identify other users. Technology can but is currently not used to be helpful in this respect.

Apart from the study by Omran and van Etten (2007), there are no other social network analyses specifically for geo-data with which we can compare our results. There are two marked differences between our work and the study by Omran and van Etten:

1. Their study was focused on understanding motivations for data sharing and how this was related to network topology. We were interested in how the network can be used for more the innocent purpose of sharing of metadata, requests for help, feedback on product quality, innovative ideas, and so on. Also, we took a greater interest in sharing across organisational boundaries, including more organisations in our analysis.
2. The people who handle the external links are different: only indirect users (Omran and van Etten) vs. mostly intermediaries and direct users (Tables 5, 6). Possibly this difference can be explained from organisational and cultural variables.

Understanding the almost inevitable non-response is an important topic in social network analysis (Brewer 2000; Marsden 2005). To our knowledge the work presented here is the first social network analysis on spatial data infrastructures. As well as empirically interesting it leads to considerations for future improvement. Possibly data acquisition can be refined by taking into account that users seem to recall at the organisational level or at the level of models rather than datasets.

Conclusions and further research

The internet revolution and evolving data infrastructures result in more efficient data sharing. Without additional efforts to strengthen feedback (user-to-user, user-to-producer and producer-to-user), these developments will result in greater anonymity and increased difficulties in finding out who is actually using a particular dataset. As a consequence sharing of feedback, concerns, ancillary and innovation among users and from users to producer may be impeded. SDI development to-date has been mostly driven by users' demand and producers efforts to make spatial data better accessible. Additional technologies (like social networking software or community websites), serving to facilitate the exchange of feedback between users have to our best knowledge not been taken-up. With lack of such support, the only thing users can do is fall back on their social network (of users of the same dataset).

We have analysed the social networks of two datasets commonly exchanged through the Dutch national SDI. We found that the majority of users has 0, 1 or 2 links to other users of the same dataset and found that there are virtually no links between users in different organisations. This seems very low and the intriguing question is why. Is it an artefact of inaccurate or incomplete response? We have presented suggestions for improvement in this paper. Are users unaware of the possible benefits of more user interaction? Or are users aware but quite happy with the current set-up? Or is the number of links low because no-one is taking the lead in promoting exchange of feedback?⁷ Understanding motivations and impediments for exchange of feedback is a major objective for further research that follows from the outcomes of our research.

The work presented in this paper can contribute to methodological research on monitoring SDI programmes (Crompvoets et al. 2007; Georgiadou et al. 2006; Harvey and Tulloch 2006; Masser 2006). Working with a clearly demarcated population (the common denominator is the use of a particular

dataset) makes the study reproducible and repeatable, thus fit for monitoring purposes. At the same time, the population definition allows individual users to exit and enter, which is in line with the dynamic nature of SDIs and organisations. Furthermore working with a larger population reduces the risk that results are affected by bias in the response of one or few informants. This is relevant in the field of data infrastructures because there may be concerns about the independence of informants (Rhind 2000). A survey such as presented here reduces the risk of bias and focuses on the actual users of data.

For further research we recommend mapping more social networks for datasets shared through data infrastructures. We hypothesise that the network topology will differ depending on cultural and organisational factors (De Vos 2007; Omran and van Etten 2007; Harvey and Tulloch 2006) and that it will differ depending on whether the data are shared either completely anonymous over the world wide web to users all over the world (e.g. as in Engler and Hall 2007) or less anonymous such as for commercial datasets and/or datasets with the user community confined to the borders of a country. We recommend research into extensions of the current method that may yield a more complete mapping of the social network. For monitoring SDI, we recommend to select a number of datasets and map their networks at regular intervals in time. And finally, we need to find out more about users' perceived benefits of user-to-user linkages and perceived obstacles to strengthening such linkages.

Acknowledgments The author would like to thank the three anonymous reviewers for their comments. This research was funded by Wageningen University, Centre for Geo-Information and by the RGI-116 project.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

The survey contained 10 questions numbered Q1 to Q10. Explanations and instructions for filling in were included in the survey but are not included in this appendix. Several things were done to get a high response to the survey.

⁷ Producers with their high number of links are in a key position to take the lead in promoting exchange of feedback. Social networking software can be used to support this. In case of the LGN dataset a user conference was organised where users could meet. Our research showed that this did not reflect in the social network. This too feeds the need for deepening our understanding of motivations and impediments for exchange of feedback.

- A short survey
- An associated website with background information
- A clear definition of whom we were interested in
- In case of non-response a reminder was sent after 1 month and if necessary a second reminder after 2 months

Several things were done to get a high response and on our request for other users' names in Q3:

- Urged to answer question Q3 as completely as possible, including also names of users in other organisations
- Strongly requested to spend sufficient time on Q3 ("all other ones are very short").
- An error message that would pop-up in case of non-response on Q3, again stressing the importance of the question for our research

Q1 Who are you and where do you work?

- Name:
- Email:
- Organisation:
- Sub organisation

Q2 What user role(s) do you have?

- Intermediary = data manager, buyer/salesperson
- Direct user = I work directly with the geo-dataset
- Indirect user = I do not work directly with the geo-dataset. But I do know the dataset and it is used in processes in which I am involved
- Ex-user = I did at one time use the dataset but not during the past year

Q3 Try to name as many other users possible, including users outside your organisation?

Q4 Do you make images/tables/graphs for "unaware" users?

- Frequently
- Sometimes
- Never

Q5 As a user, how do you give feedback to the producer?

- Not applicable, I'm an intermediary

- Have at one point in time been in touch with the data producer
- Attended a user conference
- I don't give feedback

Q6 Did you need help from someone in accessing the dataset? (C)

- Yes
- No
- Don't remember

Q7 If you were helped, who helped you?

Q8 If you want to know more about the dataset, for example about its quality or its rights, who would you ask for information?

Q9 If you want to know more about the dataset, where do you look use first?

- Metadata report
- www.lgn.nl/www.hgnnederland.nl
- Search engine (e.g. Google)
- First ask another person
- Other, namely

Q10 Do you use the original geo-dataset or do you first change it before using it?

- No changes
- Conversion to another file format
- Generalisation/aggregation
- Change projection
- Conversion to vector
- Adding own attribute data
- Changing class names but not the data themselves
- Other, namely:

References

- Albert, R., Jeong, H., & Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, 406(6794), 378–382. doi:10.1038/35019019.
- Azad, B., & Wiggins, L. (1995). Dynamics of inter-organizational geographic data sharing: A conceptual framework for research. In H. J. Onsrud & G. Rushton (Eds.), *Sharing geographic information* (pp. 22–43). New Brunswick, NJ: Centre for Urban Policy Research.
- Borgatti, P. (2002). *NetDraw: Graph visualization software*. Harvard, MA: Analytic Technologies.
- Brewer, D. D. (2000). Forgetting in the recall-based elicitation of personal and social networks. *Social Networks*, 22(1), 29–43. doi:10.1016/S0378-8733(99)00017-9.

- Budhathoki, N. R., Bruce, B., & Nedovic-Budic, Z. (2008). Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal*, 72(3–4), 149–160. doi:10.1007/s10708-008-9189-x.
- Campbell, H., & Masser, I. (1995). *GIS and organizations how effective are GIS in practice?*. London: Taylor & Francis.
- Carrington, P. J., Scott, J., & Wasserman, S. (Eds.). (2005). *Models and methods in social network analysis*. New York: Cambridge University Press.
- Clinton, W. J. (1994). Coordinating geographic data acquisition and access: The national spatial data infrastructure. Presidential executive order 12906. *Federal Register*, 59(71), 17671–17674.
- Crompvoets, J., de Bree, F., van Oort, P., Bregt, A., Wachowicz, M., Rajabifard, A., et al. (2007). Worldwide impact assessment of spatial data clearinghouses. *URISA Journal*, 19(1), 23–32.
- de Bruin, S., de Wit, A. J. W., van Oort, P. A. J., & Gorte, B. G. H. (2004). Using quadtree segmentation to support error modelling in categorical raster data. *International Journal of Geographical Information Science*, 18(2), 151–168. doi:10.1080/13658810310001629600.
- de Vos, H. J. (2007). Organisational culture: Institutionalisation of GIS for forest monitoring in Costa Rica. *Environment and Planning, B, Planning & Design*, 34(2), 355–368. doi:10.1068/b32113.
- de Wit, A. J. W., & Clevers, J. G. P. W. (2004). Efficiency and accuracy of per-field classification for operational crop mapping. *International Journal of Remote Sensing*, 25(20), 4091–4112. doi:10.1080/01431160310001619580.
- Elwood, S. (2008). Grassroots groups as stakeholders in spatial data infrastructures: Challenges and opportunities for local data development and sharing. *International Journal of Geographical Information Science*, 22(1), 71–90. doi:10.1080/13658810701348971.
- Engler, N. J., & Hall, G. B. (2007). The internet, spatial data globalization, and data use: The case of Tibet. *The Information Society*, 23(5), 345–359. doi:10.1080/0197240701572822.
- Georgiadou, P. Y., Rodriquez-Pabón, O., & Lance, K. T. (2006). SDI and e - governance: A quest for appropriate evaluation approaches. *URISA journal: journal of the Urban and Regional Information Systems Association*, 18(2), 43–55.
- Goodchild, M. F. (2008). Commentary: Whither VGI? *GeoJournal*, 72(3–4), 239–244. doi:10.1007/s10708-008-9190-4.
- Goodchild, M. F., Fu, P., & Rich, P. (2007). Sharing geographic information: An assessment of the geospatial one-stop. *Annals of the Association of American Geographers. Association of American Geographers*, 97(2), 250–266. doi:10.1111/j.1467-8306.2007.00534.x.
- Groot, R., & McLaughlin, J. (2000). *Geospatial data infrastructure: Concepts, cases and good practice*. Oxford, UK: Oxford University Press.
- Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California, Riverside (published in digital form at <http://faculty.ucr.edu/~hanneman/>).
- Hansen, M. T. (1999). The search transfer problem: The role of weak ties in sharing knowledge across organizational subunits. *Administrative Science Quarterly*, 44(1), 82–111. doi:10.2307/2667032.
- Harvey, F., & Tulloch, D. (2006). Local-government data sharing: Evaluating the foundations of spatial data infrastructures. *International Journal of Geographical Information Science*, 20(7), 743–768. doi:10.1080/13658810600661607.
- Hazeu, G. W. (2005). *Landelijk Grondgebruiksbestand Nederland (LGN5) : vervaardiging, nauwkeurigheid en gebruik*. Wageningen: Alterra, Alterra-rapport 1213, 92 p.
- Knol, W. C. (2003). *Historisch grondgebruik Nederland : tijdreeksen grondgebruik Noord-Holland van 1850 tot 1980*. Wageningen: Alterra Alterra-rapport 751, 49 p.
- Knol, W. C., Kramer, H., & Gijsbertse, H. (2004). *Historisch Grondgebruik Nederland: een landelijke reconstructie van het grondgebruik rond 1900*. Wageningen : Alterra, Alterra-rapport 573, 89 p.
- Marsden, P. V. (2005). Recent developments in network measurement. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 8–30). New York: Cambridge University Press.
- Masser, I. (2005). *GIS worlds: Creating spatial data infrastructures* (p. 312). Redlands, CA: ESRI Press.
- Masser, I. (2006). Evaluating (N)SDIs: Some general questions. Wageningen Workshop on (N)SDI impact assessment http://webdocs.dow.wur.nl/internet/grs/Workshops/AssessSDI/Masser_Wageningen%20May%202007.pdf (last accessed 16-4-2009).
- Masser, I., Rajabifard, A., & Williamson, I. (2008). Spatially enabling governments through SDI implementation. *International Journal of Geographical Information Science*, 22(1), 5–20. doi:10.1080/13658810601177751.
- Morville, P. (2005). *Ambient findability: What we find changes who we become*. Sebastopol, CA: O'Reilly Media Inc., 204.
- Nedovic-Budic, Z., & Pinto, J. (1999). Interorganizational GIS: Issues and prospects. *The Annals of Regional Science*, 33(2), 183–195. doi:10.1007/s001680050100.
- Nedovic-Budic, Z., & Pinto, J. K. (2000). Information sharing in an interorganizational GIS environment. *Environment and Planning, B, Planning & Design*, 27(3), 455–474. doi:10.1068/b2652.
- Omran, E. L. E., & van Etten, J. (2007). Spatial-data sharing: Applying social-network analysis to study individual and collective behaviour. *International Journal of Geographical Information Science*, 21(6), 699–714. doi:10.1080/13658810601135726.
- Rhind, D. (2000). Funding an NGDI. In R. Groot & J. MacLaughlin (Eds.), *Geospatial data infrastructure : Concepts, cases, and good practice* (pp. 39–56). Oxford: Oxford University Press.
- Scharl, A., & Tochtermann, K. (2007). *The geospatial web: How geobrowsers, social software and the web 2.0 are shaping the network society*. London: Springer.
- Thunnissen, H. A. M., Jaarsma, M. N., & Schouwman, O. F. (1992a). Land cover inventory in the Netherlands using remote sensing: application in a soil and groundwater vulnerability assessment system. *International Journal of Remote Sensing*, 13(9), 1693–1708. doi:10.1080/01431169208904221.

- Thunnissen, H. A. M., Olthof, R., Getz, P., & Vels, L. (1992b). *Grondgebruksdatabank van Nederland vervaardigd met behulp van Landsat Thematic Mapper opnamen*. Report 168, DLO Winand Staring Centre, Wageningen.
- van Oort, P. A. J., Bregt, A. K., De Bruin, S., De Wit, A. J. W., & Stein, A. (2004). Spatial variability in classification accuracy of agricultural crops in the Dutch national land-cover database. *International Journal of Geographical Information Science*, 18(6), 611–626. doi:[10.1080/13658810410001701969](https://doi.org/10.1080/13658810410001701969).
- VROM. (2008). *GIDEON – Basisvoorziening geo-informatie in Nederland. Visie en implementatiestrategie (2008-2011)* (p. 51). Den Haag, The Netherlands: Ministerie van VROM.
- Wehn de Montalvo, U. (2003). *Mapping the determinants of spatial data sharing*. Ashgate, UK: Aldershot.
- Williamson, I., Rajabifard, A., & Feeney, M. E. F. (Eds.). (2003). *Developing spatial data infrastructures: From concept to reality*. London, UK: Taylor & Francis.