BÁLINT GYÖRGY KUBIK AND BORÓKA PÁPAY *
The Boundaries and External Connections of the Hyperlink Network of Hungarian Websites in Romania

* [kubikbalint@gmail.com] (Hungarian Academy of Sciences Centre for Social Sciences Institute for Political Science); [papay.boroka@tk.mta.hu] (Hungarian Academy of Sciences Centre for Social Sciences "Lendület" Research Center for Educational and Network Studies (RECENS))

Abstract

The current paper describes exploratory research into the hyperlink network of Hungarian organizations in Romania. The research examines where the boundaries of the organizational system lie, and the way actors are connected to other Hungarian, Romanian and international organizations. The aim was to combine classic sociological knowledge about the Hungarian minority and its organizations with the theory of social network analysis and hyperlink analysis. The research finds that the hyperlink network of Hungarian websites in Romania is highly interconnected. Results also show that the members of the network exhibit comparably strong ties to Hungarian-language and Romanian-language websites, but the greatest proportion of external connections are with international sites. We found that the structural positions of actors within the examined network was not correlated to the distribution of external connections; actors in the network developed connections to Hungarian-language, Romanian-language and international websites regardless of their network positions.

1. Introduction

The Hungarian minority from Romania may be considered a 'society' which has its own reality (Bárdi, 1999; Kántor, 2004; Kiss, 2006). The differentiation between a minority society and a majority does not occur through everyday life interactions, but emerges through institutions (practically, organizations) (Brubaker - Feischmidt – Fox - Grancea, 2006). Thus, Hungarian organizations that act to sustain ethnicity and which satisfy communal demands that majority organizations are unable to fulfill are of elevated importance (Kiss, 2006). They are connected to both Romanian organizations and organizations from Hungary–for example, they receive various resources (including information and knowledge, in addition to funding) in different ways: from internal sources and from the majority society through other organizations (Brubaker et al., 2006; Kiss, 2006). They are also interconnected with international organizations.

The current paper discusses the results of exploratory research that examined where the boundaries of the organizational system are, and how the actors within it are connected to other Hungarian, Romanian, and international organizations. The research puts emphasis on investigating the central actors in the network, as well as their external and internal connections. There has been much theoretical discussion about the structure and connections of this organizational sphere (Bíró, 1998; Kiss, 2006), but very little empirical research. Implementing an exhaustive survey that analyzed the connections of Hungarian organizations in Romania would not only be expensive, but also hard to execute. The methodology applied here is hyperlink network analysis, where the observed hyperlink network is interpreted as a social network. This is in line with the approach of an increasing number of researchers who use hyperlink analysis to understand social phenomena and networks (Barnett et al., 2011; Heimeriks et al., 2006; Park, 2003; Park et al., 2002). Our units of analysis were the websites that operate from Romania and their Hungarian content, while we attempted to map their entire hyperlink networks using automated methods, including an internet bot developed using Python programming language and language recognition facilities. Although the hyperlink analysis of websites may seem to be far-fetched in terms of the social network analysis of Hungarian organizations from Romania, we argue that a hyperlink analysis is valid. Many hyperlink analyses have already been used to describe the physically scale-free nature and small-world properties of networks (Adamic and Adar, 2005; Albert, Jeong and Barabási, 1999; Newman, 2003). The findings of such research imply that there are influential actors in networks that support the flow of information and can help contextualize our results, showing how connected Hungarian websites in Romania really are in reality.

Automated link extraction is a good method for testing theoretical assumptions about the connectedness of organizations. The current study uses literature about Hungarian minorities and Hungarian organizations in Romania to understand the structure of the hyperlink network, and takes into account hyperlink network literature as well. Accordingly, the article is also an attempt to integrate the two approaches.

2. Hungarian society and institutions from Romania

The Hungarian minority society in Romania can be considered a complex society that has its own elite and its own organizational system (Bárdi, 1999). Although there is no consensus about the theoretical framework with which to investigate ethnic minorities, most studies treat such minorities as independent entities (Kántor, 2004).

Two hypotheses exist to describe how dependent Hungarian organizations are on the two major societies and how Hungarian minorities are related to Romanian society. The first is that the Hungarian minority creates its own reality and society within the Romanian one (Bárdi, 1999; Brubaker et al., 2006). In this framework, the institutions in the Romanian and the Hungarian community in Romania have similar functions. There are Hungarian kindergartens, schools, universities, churches, media and NGOs that make it possible for ethnic Hungarians to socialize almost completely in Hungarian, although many of these organizations, such as schools, function within the same system as their Romanian counterparts. The socialization of individuals in the 'Hungarian world' happens through an institutional system that is mainly responsible for the preservation of the minority (Brubaker et al., 2006). Kántor's framework (Kántor, 2000) describes national elites (politicians and intellectuals) as the force behind the definition and stabilization of the minority group's boundaries. The elite desires to build an independent society through the politics of creating separate institutions and creating a discourse of 'togetherness'. This institutional framework integrates not only informal relations, but organizations, alliances and policies (Brubaker et al., 2006).

The second hypothesis is that the 'Hungarian world' in Romania is an enclave that is embedded in wider society (in this example, Romanian majority society) and functions as a part of it. However, this world is also strongly connected to Hungary. At the level of the organization, it is practical and almost unavoidable that Hungarian organizations from Romania will be connected to Romanian organizations as they are part of the same administrative system. Their sources of finance include the Hungarian state (Pápay, 2014), as Hungarians from Romania often define themselves as part of the Hungarian nation (Kántor, 2004). Both hypotheses probably contain some elements of truth: the Hungarian minority exists somewhere between the 'Hungarian world' and the 'enclave' (Brubaker et al., 2006).

The most important indicator and criterion of ethnic affiliation is language (Brubaker et al., 2006). The Hungarian language is a strong glue which creates connections between Hungarian organizations (for example, Hungarian media in Romania can easily quote Hungarian news outlets). Ethnicity can be analyzed at least three levels (Barth, 1996). The micro-level focuses on the individual, and the shaping of individual identities (Barth, 1996; Hires-László, 2016). According to Brubaker (2006), the language use of individuals in everyday life is a tool of everyday communication, but not a goal. Use depends on the communication partner's language and the situation, although shared knowledge of a language is a substantial criterion of ethnic affiliation. However, individual use of the Hungarian language is not the main indicator of ethnic minority status. Reproduction of the minority occurs at the second level, which is the level of leaders, entrepreneurs and ethnic rhetoric (Barth, 1996; Hires-László, 2016). This research described in this paper takes place at

this level, since we have attempted to analyze the websites of organizations that communicate in Hungarian. The first research question is thus 'to what extent are Hungarian websites in Romania connected to Hungarian and Romanian websites?' There is also a third level of analysis, which is the level of state politics and global discourse (Barth, 1996; Hires-László, 2016).

Relevant organizational research dedicates most of its attention to NGOs (Bíró, 1998; Kiss, 2006; 2010; Kiss et al., 2004; Pápay, 2014), and their ties to state-financed organizations. It is common knowledge that the NGO sector plays a key role in the functioning of the Hungarian community in Romania. However, it has also been argued that the structure of the current institutional system has some anomalies (Kiss, 2006), and is the result of numerous historical-social processes (Bárdi, 1999; Bíró, 1998). Compared to the NGO sectors of other European societies, there are a greater number of organizations with a cultural profile (Kiss, 2010). Even the financing systems which are available reproduce a culture-centric NGO sector as the major sources of funding are the Romanian and the Hungarian state (Pápay, 2014), just as the Hungarian minority from Romania is connected to (and dependent on) both Hungarian and Romanian majority society (Kiss, 2006). According to Biró (1998), 80 or 90 per cent of Hungarian organizations from Romania have permanent ethnic goals whose priority is prestige-building in response to the majority community. Such organizations obtain resources based on the ideological considerations of external sources, not from the level at which they are active.

The current study focuses more on the structural characteristics of the network, not on the types of actors within it. At this point we make a minimum of assumptions about the structure of the sector under examination, and the connections within it.

Another key element of interest in the research is who the central actors in the network are. We assume that their connections should be differently interpreted to the ties of marginal actors. Connections with the outside word are measured through the connections of these central actors, as they are hypothesized to play a distinct role according to hyperlink analysis literature. According to the literature about Hungarian organizations from Romania, the central actors are QUANGOs, or quasi-governmental NGOs (Ágh, 1999). The Hungarian minority elite has created numerous institutions within the NGO framework to replicate the governmental institutions of the majority society and to fulfill governmental functions. However, these NGOs do not have a civic nature. Instead, they can be characterized through their relationship with majority government institutions, not by their opposition to them.

3. Hyperlink networks

Hyperlinks may be considered the basic structural element of the internet. They represent ties between websites that facilitate simple and direct contact between individuals and groups, regardless of national borders. Hyperlinks, whose function is to communicate and coordinate, are an important structural component of the World Wide Web (Park, 2003).

In the growing social science literature, hyperlink networks are interpreted as communicational systems of knowledge production and knowledge dissemination

(Heimeriks et al., 2006). Heimeriks and Besselaar (2006) interpret hyperlinks as associations between different websites, web pages, or web spheres. These ties of collaboration may emerge between users and (co-)producers of knowledge, and surface in clusters or communities. Park (2003) distinguishes between social networks and communication networks. The former consist of individuals, groups, organizations, and nation states that are in any kind of social relation with each other, while the latter are composed of individuals that are linked by information flows. Hyperlink networks can be considered a form of communication network, while nodes are websites (which may be created by individuals, groups, organizations, or nation states) whose edges are hyperlinks that represent flows of information (Park, 2003). Hyperlinks may therefore considered to be collaboration channels between actors, and communicational ties that facilitate the flow of information. There is evidence of a strong organizational incentive to form connections: bridging of social capital in this manner can increase the ability of an organization to mobilize resources (Park et al., 2002).

The analysis of a (hyperlink) network of websites shows the presence of highly connected nodes. Hubs with numerous incoming edges (a large in-degree count) suggest the existence of an organizing principle that Barabási (2016) refers to as 'scale-free property'. Barabási and Albert (1999) provide a description of a dynamic algorithm that can be used to model scale-free networks: in this framework, nodes are introduced gradually and over time, while edges are formed in a preferential manner (newly introduced nodes are more likely to develop ties to actors that are already more connected). The Barabási-Albert model can simulate networks whose degree of distribution follows a power law distribution rather than a Poisson distribution process, as suggested by earlier work about random models such as that of Erdős-Rényi (1959). Power law distribution can be formulated thus:

$$p_k \sim k^{-\gamma}$$

where k is the degree and γ is the degree exponent. For directed networks such as hyperlink networks, distribution should be approximated separately for in-degree and out-degree distribution. Barabási, Albert, and Jeong (2000) show that the topology of hyperlink networks exhibits universal scale-free properties: although the hyperlink network of websites is gigantic in size, the internet overall seems to follow scaling law processes, hinting at the existence of a very interactive, self-organized system. This study also assumes that the network of Hungarian websites in Romania is a scale-free network: we expect to find a relatively small number of vertices that are highly connected (i.e. which have a large number of incoming or outgoing edges), while the majority of nodes have a considerably smaller number of ties. Barabási and Albert (1999) emphasize how the average shortest path length (the mean of the shortest paths between every pair of vertices) changes as a function of network size in a scale-free network such as a hyperlink network. This metric is highly important since it gives an indication of how simple it is on average to find any vertex (starting from an arbitrary node) by following the intermediate edges that separate them. The authors' findings reveal that a network of websites forms a small-world network: unit growth in the size of the network does not entail linear increases in the length of the average shortest

path; the relationship is instead logarithmic. Nodes (or websites, in the case of hyperlink networks) attempt to gain access to or distribute information in the network. One efficient strategy for doing this is connecting with influential neighbors in the network so as to reach many other nodes. The tendency to connect to influential nodes amplifies the emergence of hubs and leads to improved paths of information transmission (Adamic and Adar, 2005; Adamic, Lukose, Puniyani and Huberman, 2001), resulting in a well-connected network, regardless of size. This study also quantifies how easy it is for nodes in the network of Hungarian sites in Romania to reach other sites, a metric of the level of connectedness.

A hyperlink network is formed by culture and intercultural communication, globalization theory, economics, and culture (Barnett et al., 2005). The internet is a centralized structure in terms of globalization, in which there are semi-peripheries and peripheries. There is also evidence of decentralization in the global hyperlink network. Decentralization means the emergence of regional areas and other units shaped by geographical and cultural homophily. These decentralized units exhibit independence and actively reduce the tendency to centralization (Barnett et al., 2011). Halavais (2000) claims that one can still 'see' national borders on the internet. Transnational interaction and domestic integration may be simultaneously present (Stark et al., 2005).

4. Research questions/hypotheses

Hungarian minority society in Romania has been described as a society that has its own elite and its own organizational system (Bárdi, 1999). The reproduction of the minority occurs at the 'second level'; the level of leaders, entrepreneurs and ethnic leaders, where the main actors are the institutions created by the elites. This institutional framework consists of informal relations, organizations, alliances and policies (Barth, 1996; Brubaker et al., 2006; Hires-László, 2016; Kántor, 2000). Besides organizations' websites and online media outlets, a few blogs were also included in our sample, all of which can be located at the second level of the minority community (Barth, 1996). The most important indicator and criterion of their ethnic affiliation is language (Brubaker et al., 2006). Accordingly, we relied on automated language recognition to identify our units of analysis.

Centralization and decentralization can occur on the internet at the same time when regional areas shaped by geographical and cultural homophily represent themselves as separate units (Barnett et al., 2011; Halavais, 2000; Stark et al., 2005). Based on these theories, we expect the hyperlink networks of Hungarian websites in Romania to be highly interconnected. We define interconnectedness as the ability of a site in the network of Hungarian websites in Romania to reach another website in the same network. This property is quantified through the concept of average path length and network diameter that we detail further in the description of our methodology (Barabási, 2016). We assume that the easier it is to reach any node, the more connected and more cohesive the network in question is.

H1: The hyperlink network of Hungarian websites in Romania is highly interconnected.

Earlier empirical results are used to put our findings into context. We compare numerous metrics from our network to information from other real-world networks such as hyperlink networks.

After examining the interconnectedness of the hyperlink network, the next question that should be answered is 'what ties does this network have with the outside world?' Hungarian organizations from Romania are connected to both Hungarian organizations from Hungary, Romanian entities (Kiss, 2006), and, supposedly, to other international organizations. First, to identify whether the Hungarian community in Romania can be better described as a different community to the Romanian one, or as an enclave that is perfectly embedded in Romanian society (Brubaker et al., 2006), we seek to identify whether Hungarian organizations connect to Romanian organizations at a different intensity or in a different way. We assume that Hungarian organizations from Romania are densely tied to Hungarian organizations from Hungary for many reasons. For example, they may be partly funded by the Hungarian state and state-financed organizations (Pápay, 2014), while a common national identity can also create bonds (Kántor, 2004). A common interpretation of a hyperlink network is an information (knowledge) exchange network that can also reflect collaboration (Heimeriks et al., 2006). In the case of an informational network, language can also be a powerful instrument for connecting Hungarian organizations to Hungary. Thus, we formulate the research question whether ties can be interpreted as communication paths for information exchange that favor Hungarian language partners, or if they take the form of collaborative exchanges that are strategically more advantageous to local (Romanian) partners.

RQ1: Do members of the network of Hungarian websites in Romania exhibit stronger ties to Hungarian-language than Romanian-language websites?

To explore the patterns of ties to the outside world, it is important to examine how the structural positions of the different actors affect their connections to the Hungarian, Romanian, and international sphere. As noted earlier, it appears beneficial for such actors to connect with both Romanian and Hungarian partners, but ties to international actors may also provide access to knowledge that may not be present in the local sphere. Central members may play a key role as producers and disseminators of knowledge, so the effect of centrality on external patterns of connectivity is a relevant topic for investigation.

RQ2: Do Hungarian websites in Romania with more central positions have more external connections?

5. Methodology

This chapter provides an insight into the methods of data collection and analysis which were applied. Data were collected in an automated manner and the results were manually checked to ensure their validity and to filter out irrelevant observations. Hyperlink networks may be interpreted as social networks, so the tools of social network analysis were considered applicable (Jackson, 1997).

5.1 Data collection

As noted earlier, we attempted to map the hyperlink network of Hungarian websites in Romania to reveal the network's internal structure and its connectedness to outside actors. The large number of sites necessitated the use of automated methods of analysis. Our custom-made internet bot had two main tasks: firstly, to identify the ties (hyperlinks) between relevant websites; secondly, to collect textual information from the web pages to permit determination of their language.

The unit of analysis is single websites. Manual inspection of the sample revealed that the majority of websites in our sample could safely be considered to belong to organizations, so we made the decision to refer to these websites as organizations. As noted earlier, three levels of an ethnic society can be distinguished, the second level being where the reproduction of society happens (Barth, 1996; Brubaker et al., 2006). The few personal websites in our database belong to the second level of ethnic society.

It is important here to provide an insight into the technical details of the process of data collection. A website contains a number of hyperlinks that point either to its own subpages, or other websites. The research interest in this case is in the latter: collecting the hyperlinks to different sites allows us to map the hyperlink network of interest. Websites also contain textual information that needs to be processed to identify the language. Making such a language classification was crucial to addressing the hypotheses and research questions. A custom-made internet bot was created to undertake this task. The computer program was developed in Python, and is able to process a large number of websites in parallel and thus collect data in a short period of time. The bot opens the websites of interest, collects hyperlinks from the main page and (up to a hundred) subpages, and analyzes site-related text to determine its language. We chose a standalone language identification tool named langid.py[1] that has been pre-trained on numerous languages, including Hungarian and Romanian.

The following paragraphs describe the method of sample building. The web bot started its life cycle by processing a starting sample that was collected from a public database[2] of Hungarian websites in Romania compiled by Transindex[3]. The next step was to analyze the hyperlinks collected from sites in the starting sample. Texts from every website were collected and processed to determine their language, but hyperlinks were only collected from sites with Romanian country code, top-level domains (.ro) with content written in Hungarian. Inaccessible sites were not included in the sample. Subsequent rounds of data collection followed the same logic. A total of six cycles were executed until the data saturation point was reached (the sixth iteration collected data about largely redundant websites). We also manually filtered out websites that were most likely members of the Hungarian hyperlink network in Romania and identified pages that were not present in the starting sample. It seems

---

[1] For more information on the project's GitHub repository: https://github.com/saffsd/langid.py Accessed: 21-03-2017.

[2] The database can be found at: http://netkatalogus.adatbank.transindex.ro. Accessed: 21-03-2017.

[3] It should be noted here that the Transindex sample also contained websites with non-Romanian, top-level domains (.hu, .com, etc.).

safe to assume that the overwhelming majority of Hungarian websites in Romania have been mapped and the sample sufficiently represents the whole population.

We made the decision to exclude certain domain names from the database. Websites connected to the most popular search engines (such as Google.ro), social networking sites (such as Facebook.com), and blog services (for example, blogspot.ro) were not parsed to prevent the inclusion of a large number of potentially irrelevant sites. Inclusion of these sites in the database would have resulted in more analysis-related challenges and necessitated the additional step of a costly manual filtering process (this exclusion did not apply to the first cycle of web bot activity).

5.2 Limitations of the data

One limitation of the database is the selection mechanism for the websites, as sites were selected mainly through identification of the country-level domain. The first web crawl collected all the websites included in the Transindex sample, but later cycles only identified websites as members of a network with Romanian top-level domains (.ro) and Hungarian textual content. This sampling criterion may have led to the exclusion of sites that had other country-level domains (.hu for example), but which are in fact part of the network of Hungarian websites in Romania. This limitation is somewhat offset by the fact that the first cycle of data collection included all sites in the starting sample, regardless of their domains. The same logic applies to the identification of external partners. There may be websites that use Romanian domains and contain Hungarian content, but which are not maintained from Romania; the same may be true in the case of Hungary. One reason for this is the presence of diasporas.

Another important limitation is that we did not collect data for the Romanian hyperlink network as a whole. Analysis of Hungarian websites in the context of the entire Romanian world wide web would have been advantageous (allowing us, for example, to identify whether the network of Hungarian websites in Romania is a clearly defined community in terms of the whole Romanian hyperlink network), but we did not have the resources to map the entire hyperlink network. We therefore made the decision to limit our attention to Hungarian websites.

5.3 Database

Using the automated method detailed earlier, we were able to map 1091 websites and 97858 hyperlink connections between members of the network of Hungarian websites in Romania. Since such hyperlinks navigate the visitor from one site to another, we consider this network to be a directed network and analyze it accordingly. After filtering out multiple edges, 2220 unique connections remained. We decided to use a simplified network approach (multiple edges and loops removed) because the amount of redundant hyperlinks was highly variable and we were more interested in the existence (or lack) of ties between websites. Figure 1 illustrates the visualized network of websites within the Hungarian network in Romania. The nodes exhibited an average of 2.04 connections to other websites in the network.
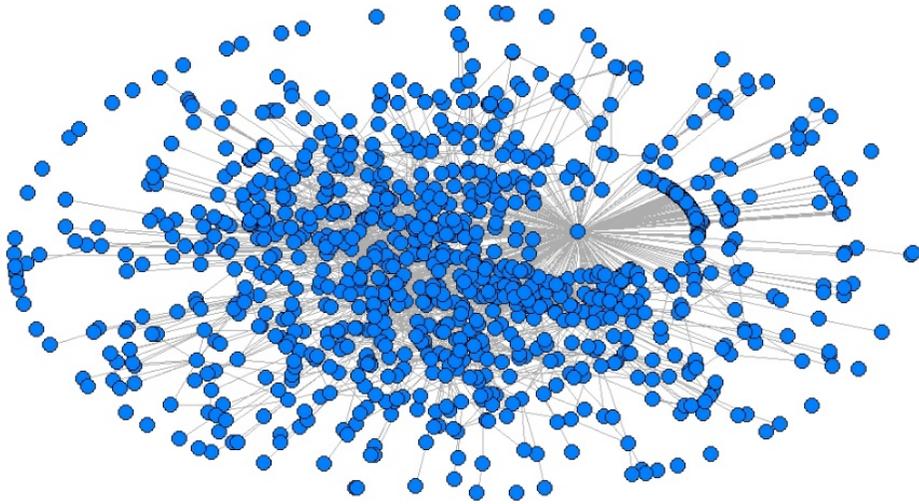
Figure 1: Visualization of the hyperlink network of Hungarian websites in Romania

The external connections of the Hungarian websites in Romania were reduced to 2434 unique websites that were functioning at the time of the data collection. The total number of ties amounted to 147719 and the removal of multiple ties resulted in 3822 unique connections. These connections were exhibited by a total of 986 Hungarian websites in Romania that had 1.89 connection on average to external partners.

The network of Hungarian websites in Romania included a large number of different actors, including numerous organizations, media outlets, cultural foundations, universities and many others. Table 1 lists the actors with the largest number of incoming hyperlink ties. As mentioned earlier, we consider the hyperlink network of Hungarian websites in Romania a scale-free network, so these sites can be safely referred to as hubs (the most connected vertices in the network). The website *hargitamegye.ro* belongs to the local government of Hargita County in Romania which has a mostly Hungarian population and uses Hungarian as its administrative language. *Rmdsz.ro* is the website of RMDSZ (Democratic Alliance of Hungarians in Romania), the largest political party for the Hungarian minority from Romania. *Communitas.ro* belongs to Communitas, a QUANGO created and sustained by RMDSZ whose goal is to allocate state funds to the Hungarian minority and related organizations. The other websites on the list (*transindex.ro, erdely.ma, szekelyhon.ro, kronika.ro, 3szek.ro, maszol.ro*) belong to media outlets.

Table 1: Nine websites in the network of Hungarian websites in Romania with the greatest number of incoming hyperlink connections

Website	Number of incoming connections (in-degree)
transindex.ro	74
communitas.ro	65
hargitamegye.ro	42
erdely.ma	41
szekelyhon.ro	39
rmdsz.ro	31
kronika.ro	31
3szek.ro	29
maszol.ro	20

It is important to note that the Hungarian websites in Romania notably prefer to develop ties to other members of the network, as opposed to ties with external partners. There is a greater chance of a website in our sample connecting with another Hungarian website in Romania than another random website. The average proportion of external partners amongst all connections is 37.6 per cent, and the median 33.3 per cent.

We identify 47 weak components in the sample network. An overwhelming majority of the nodes, however, are part of a single larger one consisting of 966 websites, while the other components usually have 2 websites. Manual inspection revealed that we can safely assume that the numerous small components are mostly due to the presence of websites containing highly outdated content and plain text. These websites were part of the initial sample provided by Transindex.hu, but eventually proved to be a dead end in terms of data due to their old content, references to dysfunctional sites, or lack of hyperlinks to other sites.

5.4 Concepts

In this study we use the term interconnectedness as a measure of how simple it is for a visitor who starts out at one of the websites in the sample to reach any other page in the network of Hungarian websites in Romania. We therefore use the concept of connectedness in a different way to Barabási, who defines connectedness as the existence of a path between two arbitrary nodes in a network (2016). The present researchers were more interested in the measure of distance between any two nodes. In networks, distance can be measured using path length, which refers to the routes that follow the edges of a network, and is the number of links of which a path consists. The average shortest path is the arithmetic mean, while the network diameter is the maximum of the shortest path lengths between any pair of network vertices. We use average path length and diameter to assess the level of connectedness of our network (Barabási, 2016).

6. Results
6.1 Interconnectedness

First, consider our first hypothesis: *the hyperlink network of Hungarian websites in Romania is highly interconnected.* To examine the truth of this statement, we investigate the connectedness of the websites in the sample.

Before discussing network measures, it is important to note that we identified 47 weak components in the sample network. An overwhelming majority of the nodes, however, are part of a single larger one consisting of 966 websites, while the other components usually have 2 websites. Manual inspection revealed that we can safely assume that the numerous small components involve websites containing highly outdated content and plain text. The websites that belong to the small components were part of the initial sample provided by Transindex.hu, but eventually proved to be dead ends in the process of data collection due to their old content, references to dysfunctional sites, or lack of any hyperlinks pointing to other sites.

The degree distribution of the network of Hungarian websites in Romania follows a power law. Since it is a directed network, both in-degree and out-degree distribution are analyzed. As noted earlier, scale-free networks exhibit degree distributions that can be well approximated by a power law. The exponent of the power law fit for the in-degree distribution is 2.578, while the fitted exponent for out-degree distribution is 2.14. The p-value for a Kolmogorov-Smirnov test was sufficiently high, affirming that the data could have been drawn from the fitted power law distribution. Appendix 1 includes numerous statistics from real-life networks, besides the network of Hungarian websites in Romania. The two other hyperlink networks included in the table also exhibited power law exponents in the range of 2-3, but in their case the exponent of in-degree distribution (2.1) was smaller than the exponent of out-degree distribution (2.4 and 2.7). We conclude therefore, that the network of Hungarian websites in Romania is a scale-free network.

As discussed earlier, the average path length and diameter of the network are considered the most important metrics for quantifying the interconnectedness of the network of Hungarian websites in Romania. Figure 2 is a histogram displaying shortest

path lengths for every pair of nodes. While the maximum length of the shortest path between any two nodes (the network diagram) is 21, the average shortest path length is 6.71 and the mode of these distances is five. We conclude that, on average, a site in the network of Hungarian websites in Romania can be reached from any other page by following a path that leads through five or six other websites.
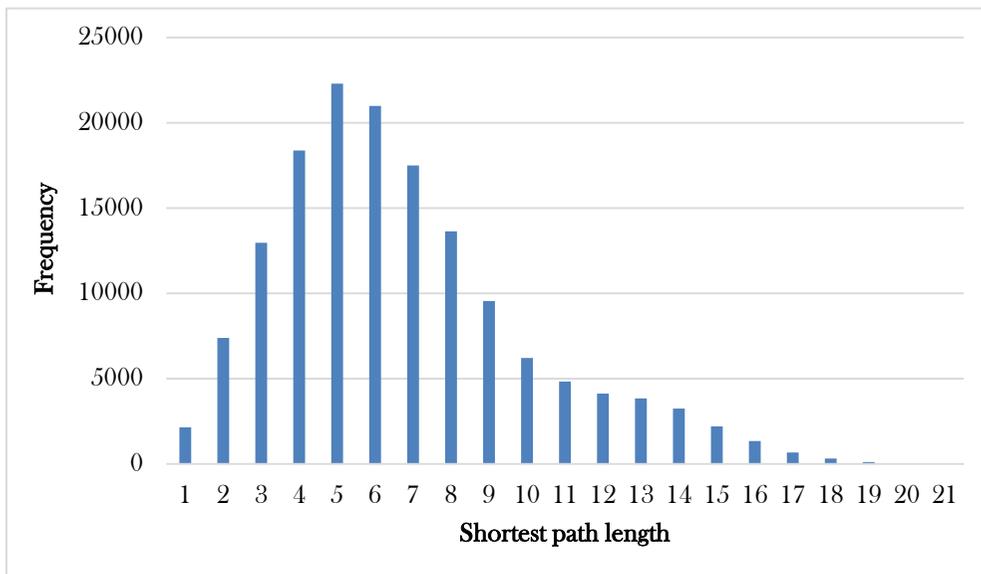

Figure 2: Shortest path lengths between any two nodes in the network of Hungarian websites in Romania

This is considerably less than the average shortest distance measured in the case of nd.edu (11.27) and the Altavista hyperlink network (16.18). This indicates the considerable interconnectedness of the network in question, compared to the hyperlink network of the internet.

The research thus finds that the hyperlink network of Hungarian websites in Romania is highly interconnected: the numerous actors in the examined network are connected as a 'society'.

6.2 Ties to the Hungarian, Romanian and international spheres

We have already seen that the network of Hungarian websites in Romania is quite interconnected, but have also identified a strong preference for connections with other members of the network, as opposed to external partners (the average proportion of external partners for the websites in the network of interest is 37.6 per cent). This subchapter provides a summary of results concerning ties to the Hungarian, Romanian, and international sphere.

We begin with a discussion of the first research question (*RQ1: Do members of the network of Hungarian websites in Romania exhibit stronger ties to Hungarian-language than Romanian-language websites?*). As mentioned earlier, we used an automated language detection process to identify the languages used on the websites in

the sample, including sites that are not part of the network of Hungarian websites in Romania. We now turn to examining the languages we were able to identify on the websites of external partners. Here, we refer to languages that are not Hungarian or Romanian as 'other'. This group overwhelmingly consists of major international languages such as English or German, but less commonly encountered languages such as Slovakian, Dutch, and Swedish also occur sporadically. We classify websites with content written neither in Hungarian nor in Romanian as 'international websites'.

Table 2: Languages used on external websites

Language	Frequency	Per cent
Other	1016	41.7%
Hungarian	711	29.2%
Romanian	707	29.0%
Total	2434	100%

As can be seen from Table 2, slightly more websites (711) used the Hungarian language than had content written in Romanian (707). It is also apparent that most external partner websites used other languages (we remind the reader of one of the main limitations of this study: that the Romanian hyperlink network as a whole was not mapped out). The one-sample Chi-square statistic is 77.454, which is statistically significant at $p < 0.99$. We therefore state that most websites that are externally connected to Hungarian sites in Romania are international, and the number of actors with Hungarian and Romanian content is approximately equal. Thus, we cannot detect a significant preference for Hungarian sites as opposed to Romanian ones, but international partners represent the majority (41.7 per cent) of external agents. We argued earlier that Hungarian organizations in Romania function within the same system as their Romanian counterparts (Brubaker et al., 2006) and they receive financial contributions from both Romanian and Hungarian sources, but their language and national identity should tie them to Hungarian-speaking websites. The aforementioned reasons for the distribution of connections to Romanian and Hungarian sites may balance each other out.

As suggested by the second research question (*RQ2: Do Hungarian websites in Romania with more central positions have more external connections?*), we also attempted to identify the presence of relationships in the structural positions of members of the network of Hungarian websites in Romania. As noted earlier, in-degree centralities are a viable indicator of this in our case, since the incoming ties for each node are assumed to provide a good estimate of structural positions in the

network. We attempted to estimate the strength of the relationship between structural positions (mainly in-degree centralities) and the proportion of partners, but could not identify a significant effect. The Pearson-correlation between the proportion of Hungarian-language partners and (normalized) in-degree centrality was 0.0303, which is very close to zero. We were also unable to identify a significant and strong correlation between structural position and the proportion of Romanian (0.0182) or international partners (-0.0404).

These results indicate that the members of the network exhibit stronger ties to international external partners than to either Hungarian or Romanian-language websites. External partners with Hungarian content do not seem to be favored above sites with Romanian text. Results also show that the structural position in the network of Hungarian websites in Romania does not affect the pattern of connectivity with the Hungarian, Romanian, or international sphere.

7. Further research

The use of social network analysis in this study of the network of actors in a minority society proved to be a useful approach. Due to the multitude of different types of ties between Hungarian organizations in Romania, we argue for the broader utilization of the technique to gain deeper understanding of the interdependence between these actors, and to better identify communities in the network.

The hyperlink network of Hungarian websites in Romania offers many opportunities for future research. As noted earlier, the decentralized nature of the network should be studied in-depth, and more detailed research is needed to distinguish between different kinds of actors. The network database could be expanded by including variables that can be used to break down the websites into organizational and/or geographical categories. This kind of classification, if factors such as the flow of monetary resources were controlled for, would enable researchers to explain the clustering patterns observed in the network. Mapping of the Romanian and Hungarian hyperlink network would permit the identification of the real positions in the network of Hungarian websites in Romania in a broader context.

The data collection process for this research involved the extraction of textual information from websites, and the corpus that was created makes quantitative text analysis possible. It would be beneficial to study phenomena such as emerging (political, geographical, etc.) topics in the network.

It is worth mentioning that the Transylvanian Hungarian community may exhibit regional differences. While in 'Szeklerland' there is a separate Hungarian reality (i.e., one distinct from Romanian society, or, in other words, a small Hungarian community within the Romanian state), the use of language, employment policy and the financial maintenance of Hungarian institutions are more problematic in the ethnically mixed territories (Bárdi, 1999). It would also be interesting to examine the spatial fragmentation that is possibly present in the network.

Although we were able to map the ties between the relevant websites, the nature of these connections leaves room for further research. It might be beneficial to classify these ties based on the context in which they have emerged. For example, an analysis

of the textual context of the hyperlinks would make it possible to identify partnerships or financial ties, which information could be used in further analyses.

8. Conclusion and discussion

The paper has described exploratory research that mapped the network of Hungarian organizations in Romania using hyperlink network analysis.

The social scientific literature about the Hungarian minority in Romania tends to distinguish three levels (micro-meso-macro) of minority society, where the second level is responsible for the reproduction of society itself. The leaders of the minority create institutions that are independent alternatives to the institutions of the majority, leading to the emergence of a community with a separate reality. In the case of the Hungarian minority in Romania, the Hungarian language is of central importance in creating and sustaining the identities of members. Even though research has examined numerous Hungarian minority organizations, only a few studies have mapped the networks of these entities.

We argue for the utilization of hyperlink analysis in the study of the network of Hungarian organizations in Romania. Hyperlink connections are considered communication-based paths of knowledge production and dissemination, and there are organizational incentives for revealing these ties. The tools of social network analysis are applicable in the analysis of hyperlink networks. The rapid growth of the international hyperlink network also shows signs of decentralization which has resulted in the emergence of national, regional, and other borders on the internet. We assume that pages of the Hungarian minority in Romania are components of such a decentralized community.

We employed automated methods to map the network of Hungarian websites in Romania and to reveal the ties of its members to the Hungarian, Romanian, and international spheres using an automatic language recognition process. The sampling method may have resulted in some instances of misclassification, which is one limitation of the research described in this paper. The database went through a process of manual inspection which validated the reliability of the data by helping to filter out irrelevant observations. Our units of analysis were websites in Romania (not necessarily those with a Romanian top-level domain) with Hungarian-language content. We mapped the network using a starting sample provided by Transindex.hu, and then by traversing hyperlink connections. A further research limitation is that the whole Romanian hyperlink network was not mapped.

We were able to show that the network of Hungarian websites in Romania is strongly interconnected, forming a community with a separate reality. When choosing external partners, websites from Hungary do not seem to be preferred to sites with Romanian-language content. Thus, there appears to be a balance between the factors that incentivize Hungarian minority organizations to develop ties to either the Hungarian or the Romanian sphere. There is, however, a strong preference for international (neither Hungarian nor Romanian) partners. The research was unable to demonstrate any relationship between structural positions in the network and patterns of connectivity with external partners.

Further research could segment the organizations in the network of Hungarian minority organizations into categories. Since we could not distinguish between different types of actors in the network, only macro-level analysis was possible. A more fine-grained analysis would involve controlling for, among many other factors, organizational profiles, geographical regions, or sources of financing.

Analyzing the hyperlinks of Hungarian websites in Romania may seem to be an approach that is positioned outside the classic theory of minority studies, but the two approaches can be integrated. Hyperlink network analysis, according to which hyperlink connections are considered communication-based paths of knowledge dissemination, can stand on its own. The approach provides researchers with an opportunity to reveal relationships that may be more complex and costly to discover using other, more traditional methods. The opportunity it offers researchers to examine the network of Hungarian entities from Romania, while taking into consideration the literature on Hungarian minorities, should not be missed.

References

Adamic, L. A., Lukose, R. M., Puniyani, A. R. and B. A. Huberman (2001) Search in Power-Law Networks. *Physical Review E,* 64(4): 046135. DOI: https://doi.org/10.1103/physreve.64.046135

Adamic, L. and E. Adar (2005) How to Search a Social Network. *Social Networks,* 27(3): 187-203. DOI: https://doi.org/10.1016/j.socnet.2005.01.007

Ágh, A. (1999) Civil társadalom és korai konszolidáció az EU-csatlakozás jegyében Magyarországon (Civil Society and the Early Consolidation in Hungary in the Light of EU Accession). In Csefkó, F. and Cs. Horváth (Eds.) *Magyar és európai civil társadalom (Hungarian and European Civil Society)*. Pécs: Pécs-Baranyai Értelmiségi Egyesület. 59-61.

Albert, R., Jeong, H., and A-L. Barabási (1999) Internet: Diameter of the World-Wide Web. *Nature,* 401(6749): 130-131. DOI: https://doi.org/10.1038/43601

Barabási, A.-L. (2016) *Network Science.* Cambridge: Cambridge University Press.

Barabási, A.-L., Albert, R., and H. Jeong (2000) Scale-free Characteristics of Random Networks: the topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications,* 281(1-4): 69-77. DOI: https://doi.org/10.1016/s0378-4371(00)00018-2

Bárdi, N. (1999) Törésvonalak keresése és összehasonlítása a határon túli magyar politikában (Finding and Comparing Cleavages in Hungarian Transborder Politics). In Székely, I. (Ed.) *Útközben. Pillanatképek az erdélyi magyar politika reformjáról (In Transit. Snapshots about the Reforms of Transylvanian Hungarian Politics).* Csíkszereda: Pro-Print Könyvkiadó. 19-44.

Barnett, G. A., and E. Sung (2005) Culture and the Structure of the International Hyperlink Network. *Journal of Computer-Mediated Communication,* 11(1): 217-238. DOI: https://doi.org/10.1111/j.1083-6101.2006.tb00311.x

Barnett, G. A., Chung, C. J. and H. W. Park (2011) Uncovering Transnational Hyperlink Patterns and Web-Mediated Contents: A New Approach Based on Cracking.com Domain. *Social Science Computer Review*, 29(3): 369-384. DOI: https://doi.org/10.1177/0894439310382519

Barth, F. (1996) Régi és új problémák az etnicitás elemzésében (Old and New Challenges in Analyzing Ethnicity). *Régió*, 7(1): 2-25.

Bíró, Z. (1998) Intézményesedési folyamatok a romániai magyar társadalomban 1989-1995 között (Institutionalization Processes in the Romanian Hungarian Society from 1989 to 1995). In Biró, Z. (Ed.) *Stratégiák vagy kényszerpályák? Tanulmányok a romániai magyar társadalomról (Strategies or Forced Paths? Essays on the Romanian Hungarian Society)*. Csíkszereda: Pro-Print Könyvkiadó. 15-48.

Brubaker, R., Feischmidt, M., Fox, J. and L. Grancea (2006) *Nationalist Politics and Everyday Ethnicity in a Transylvanian Town*. Princeton: Princeton University Press.

Erdős, P., and A. Rényi (1959) On random graphs, I. *Publicationes Mathematicae Debrecen*, (6): 290-297.

Halavais, A. (2000) National Borders on the World Wide Web. *New Media and Society*, 2(1): 7-28. DOI: https://doi.org/10.1177/14614440022225689

Heimeriks, G., and P. Van den Besselaar (2006) Analyzing Hyperlinks Networks: The Meaning of Hyperlink Based Indicators of Knowledge Production. *Cybermetrics*, 10(1): 1-9.

Hires-László, K. (2016) "Magyar világ" - magyar pillérek és oszlopok Beregszászon ("Hungarian World" - Hungarian Pillars and Columns in Beregszász). In V. Gazdag, Z. Karmacsi and E. Tóth (Eds.) *Értékek és kihívások. Vol. II. (Values and Challenges. Vol. II.)*. Ungvár: Autdor-Shark.

Jackson, M. H. (1997) Assessing the Structure of Communication on the World Wide Web. *Journal of Computer-Mediated Communication*, 3(1). DOI: https://doi.org/10.1111/j.1083-6101.1997.tb00063.x

Kántor, Z. (2000) Kisebbségi nemzetépítés – a romániai magyarság, mint nemzetépítő kisebbség (Minority Nation-Building - Romanian Hungarians as a Nationalizing Minority). *Regio*, 2000(3).

Kántor, Z. (2004) Nemzet és intézményesülés (Nation and Institutionalization). In Fedinec, Cs. (Ed.) *Nemzet a társadalomban (Nation in Society)*. Budapest: Teleki László Alapítvány. 275-290.

Kiss, D. (2006) Az erdélyi magyar civil szféráról (On Hungarian Civil Sphere in Transylvania). *Civil szemle*, 2006(2).

Kiss, D. (2010) *Romániai magyar non-profit szervezetek – 2009-2010. A szervezetek adatbázisának bemutatása és a non-profit szektor szociológiai elemzése (Hungarian Non-Profit Organizations in Romania - 2009-2010. Presenting the Database of Organizations and the Sociological Analysis of the Non-Profit Sector).* Kolozsvár: Nemzeti Kisebbségkutató Intézet.

Kiss, T., Csata, Z., and D. Kiss (2004) Az erdélyi magyar kulturális intézményrendszerről (On the Hungarian Cultural Institution System in Transylvania). *Erdélyi Társadalom*, (1): 133-162.

Newman, M. E. (2003) The Structure and Function of Complex Networks. *SIAM review,* 45(2): 167-256. DOI: https://doi.org/10.1137/s003614450342480

Pápay, B. (2014) Állami forráselosztás a romániai magyar nonprofit szektorban (The Distribution of Governmental Resources in the Hungarian Non-Profit Sector in Romania). In Bodó, B. (Ed.) *Erdélyi Magyar Civil Évkönyv (Transylvanian Hungarian Civil Annales).* Cluj Napoca: Magyar Civil Szervezetek Erdélyi Szövetsége. 82-97.

Park, H. W. (2003) Hyperlink Network Analysis: A New Method for the Study of Social Structure on the Web. *Connections*, 25(1): 49-61.

Park, H. W., Barnett, G. A., and I. Y. Nam (2002) Interorganizational Hyperlink Networks Among Websites in South Korea. *Networks and Communication Studies*, 16(3-4): 155-174.

Stark, D., Vedres, B., and L. Bruszt (2005) Global Links, Local Roots? Varieties of Transnationalization and Forms of Civic Integration. *Center for Organizational Innovation Working Papers*, 2005(4).

Appendix 1 – Network parameters

	Network	Type	Number of vertices	Total number of edges	Mean degree	Mean vertex distance	Exponent a of degree distribution if the distribution follows a power law	Transitivity	Clustering coefficient	Degree correlation coefficient	Citations for network in bibliography
social	telephone call graph	undirected	47000000	80 000 000	3.16		2.1				8. 9
social	email messages	directed	59912	86 300	1.44	4.95	1.5/2.0		0.16		136
information	WWW nd.edu	directed	269504	1497135	5.55	11.27	2.1/2.4	0.11	0.29	-0.067	14. 34
information	WWW Altavista	directed	203549046	2 130 000 000	10.46	16.18	2.1/2.7				74
information	citation network	directed	783339	6 716198	8.57		3.0/-				351
information	Roget's Thesaurus	directed	1022	5 103	4.99	4.87	-	0.13	0.15	0.157	244
information	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119. 157
information	**Hyperlink network of Hungarian websites in Romania**	**directed**	**1091**	**2220**	**4.08**	**6.73**	**2.58/ 2.14**	**0.0854**	**0.2672**	**-0.036**	
technological	Internet	undirected	10 697	31992	5.98	3.31	**2.5**	0.035	0.39	-0.189	86. 148
technological	power grid	undirected	4 941	6 594	2.67	18.99	-	0.10	0.080	-0.003	416
technological	software packages	directed	1439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	-0.016	318
technological	software classes	directed	1377	2 213	1.61	1.51	-	0.033	0.012	-0.119	395
technological	peer-to-peer network	undirected	880	1296	1.47	4.28	2.1	0.012	0.011	-0.366	6. 354