Computers, Environment and Urban Systems 85 (2021) 101565

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Information diffusion between Dutch cities: Revisiting Zipf and Pred using a computational social science approach

Antoine Peris[a,*], Evert Meijers[a,b], Maarten van Ham[a,c]

[a] Department of Urbanism, Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, the Netherlands
[b] Department of Human Geography and Spatial Planning, Faculty of Geosciences, Utrecht University, Utrecht, the Netherlands
[c] School of Geography and Sustainable Development, University of St Andrews, St Andrews, Scotland, United Kingdom

ARTICLE INFO

Keywords:
System of cities
Information flows
Gravity model
Netherlands
Historical newspapers

ABSTRACT

News travels fast and far, and the general idea is that the spatial extent of news coverage has increased over time. Information flows are always involved in systems of interdependent cities. This is the reason why George Zipf and Allan Pred, both pioneers of the urban systems literature, were eager to obtain data on these relations to understand urban system dynamics. However, because of limited resources in data acquisition, they restricted their studies to small samples of cities or short periods of time. By using novel computational social science techniques on a digital archive of historical newspapers, we could map and explore changes in the spatial extent of news coverage in the Netherlands at an unprecedented detailed scale for a period of 62 years. In this paper, we analyse 24 million news items mentioning 312 different cities and towns in a sample of 31 local newspapers. Thanks to this data, we were able to reconstruct the information field of urban readerships from different cities and how it changed over time. By analysing their evolution, we find evidence of space-time contraction with an increasing coverage of faraway places in the period ranging from 1869 to 1930. However, this coverage is not evenly distributed but is characterized by a hierarchical selection process. Coverage of the largest cities in the Randstad increased at the expense of information flows from intermediate provincial cities. More generally, this paper shows how computational social science approaches may offer new ways of looking at urban dynamics with large text corpora such as digital archives of historical newspapers.

1. Introduction

When studying the organisation of systems of cities, research can build on many different types of empirical data: the circulation of people, goods, capital or information. Among these elements, information has a particular place in the research tradition on urban systems because it plays a crucial role in organising the complex patterns of networks and flows connecting cities. The tradition to focus on information started with early research by the pioneers of the system of cities literature analysing the content of local newspapers (Pred, 1973, 1977; Zipf, 1946a) and is still active today because of the availability of new data sources on human communications through mobile devices and social media (Grauwin et al., 2017; Krings, Calabrese, Ratti, & Blondel, 2009; Stephens & Poorthuis, 2015).

Because information is a prerequisite to any other kind of exchanges, mapping its patterns helps to understand the organisation of the system of cities itself. This idea has been corroborated recently by researchers working on the development of the postal road network in France between the 17th and 19th century (Bretagnolle & Franc, 2017). They have shown that the development of an integrated communication network was concomitant with the development of the city-system at the national scale and the take-off in urbanisation rate. Having longitudinal data is very important in order to understand cities because the roots of their growth or decline are often situated far in the past. Inertia and initial advantages are indeed often major explaining factors of their situation in the urban hierarchy. However, collecting data on their relations, and especially information circulation remains a challenge because of the change in communication technology.

Some of the earliest research on urban systems has collected data from newspapers in order to analyse the circulation of information (Zipf, 1946a), because it was one of the only ways to develop a relational approach of cities at this time. But because of the cost of the data collection – involving analysing hundreds of actual paper newspapers – these studies were limited to a small number of cities, and very short

* Corresponding author at: Julianalaan 134, 2628 BL Delft, the Netherlands.
E-mail address: a.f.t.peris@tudelft.nl (A. Peris).

https://doi.org/10.1016/j.compenvurbsys.2020.101565
Received 2 November 2019; Received in revised form 12 October 2020; Accepted 26 October 2020
Available online 11 November 2020

periods of time. Today, thanks to the effort of digitization of historical newspapers undertaken by national libraries and the developments in data science, it is possible to upscale such analysis and apply distant reading techniques to extract similar data from millions of digitalized newspaper sheets. Digital humanities scholars have shown that these massive digital archives can be used to identify macroscopic trends related to cultural changes (Bod, 2013; Lansdall-Welfare et al., 2017), but very few studies have looked systematically at the geographical dimension of these archives.

In this paper, we look at the evolution of the circulation of information within the Dutch urban system with data extracted from Delpher, the digital archive of historical newspapers from the National Library of the Netherlands. This archive contains around 12 million newspapers pages containing even a much larger number of news items. We have selected a corpus of 31 local newspapers and collected the news referring to 312 settlements for the period 1869–1930. This represents a total number of 24 million of news items, which we use to analyse the evolution of the Dutch urban system over a long time period through its patterns of information flows. The overarching goal of the paper is to test the potential of this novel data source for reconstructing the evolving urban geography of an entire country. In order to achieve this goal, we will look for regularities in how flows of information develop over time and test different hypotheses related to the evolution of systems of cities.

Following our theoretical review (section 2), a variety of hypotheses will be formulated addressing the role of space and cities in information circulation. After a presentation of the novel data we use (section 3), and our initial research approach employing gravity modelling to understand the circulation of information over time (section 4), we test our hypotheses in the following section (section 5). Further explorations of our hypotheses using alternative approaches that handle spatial heterogeneity in information circulation are presented in section 6. Finally, we conclude in section 7.

2. Theory and background

2.1. Information flows in systems of cities

One of the main characteristics of cities is their ability to organize territories by articulating many types of networks and flows. Cities do not function in isolation but in systems at different scales from the regional to the global (Pumain, 2011). While recent research indicates that these relations are increasingly important (Meijers, Burger, & Hoogerbrugge, 2016), it has been shown by historians that relations were central in the fate and fortune of cities long before the current globalization (Hohenberg & Lees, 2009). For all these reasons, networks and flows are at the centre of the understanding of cities (Batty, 2013).

Information flows have long been recognized by geographers and urban scholars as an essential explanation of urban and spatial processes (Hägerstrand, 1967; Meier, 1962; Törnqvist, 1968, 1970). Researchers acknowledged very early that information was the most central resource in an urban system, because “without information about the economic risks and opportunities of the system, there would be no directed movement within the system” (Zipf, 1946a). The same idea was also emphasised in the work of Meier (1962), arguing that information and knowledge, that are conveyed by communications, were at the root of the mechanisms making economic growth possible in cities. Few years later, the empirical investigations of Allan Pred on the urban system of the United-States departed from the same hypothesis: “None of the economic actions and location decisions that underlie individual and collective urban growth can materialize unless preceded by information acquisition. None of the interurban commodity, capital, and human flows that are the outward expression of growth can transpire unless there is either the transmission of knowledge about demand, prices, and opportunities or some other form of information exchange.” (Pred, 1973, p. 2).

The connections between information and other types of circulations is the reason why it was a major focus in quantitative geography research from the 1950s onward. The dissertation of Hägerstrand, initially published in Swedish in 1953 and translated into English later on by Pred (Hägerstrand, 1967), played a big role in putting the study of information patterns at the core of the research agenda of human geography. Hägerstrand’s contribution was to propose models describing the adoption of innovations in the population of a Swedish region by means of information dissemination through personal contacts. The notion of information field was presented for the first time in his work as an operationalisation of the probability of contact between individuals. In this seminal work focusing on a rural population, this probability was decreasing homogeneously with distance. However, Hägerstrand noted that in the case of a system of cities, the urban hierarchy would channel the course of diffusion. This element was empirically observed by Pred in his analysis of the system of cities of the United-States in the 19th century (Pred, 1973), focusing on the transmission of many innovations and identifying spatial biases in the availability of information and their pivotal role in explaining the diffusion patterns. In this study, he observed the prevalence of hierarchical diffusion processes but observed also some “neighbourhood effect” with small cities in the vicinity of big ones receiving innovations more early than would be expected given their position in the hierarchy – an early example of what would become known as ‘borrowed size’ (see Meijers et al., 2016).

While the previously mentioned works were mostly conceptual or limited to small samples of cities (19 in the case of Pred, 30 for Zipf), a more recent systematic exploration at the scale of an entire urban system confirmed the intricate interweaving of communication and urbanisation (Bretagnolle & Franc, 2017). Indeed, in their study they proved that in France, the emergence of an integrated national urban system and the take-off of urbanisation were concomitant with the development of a large scale communication network of postal roads around the middle of the 18th century.

The aforementioned body of literature had influence beyond the Swedish and North-American quantitative geography. In the 1990s works from urban sociologists on the global cities (Sassen, 1991) and the networked society (Castells, 1996) were influenced by the conception of cities as centres processing information inherited from the urban literature of the 1950s-1970s. The empirical counterpart of this emerging research tradition, and especially the work of the Globalization and World Cities research group (Derudder et al., 2003; Taylor, 2001), is explicitly focusing on the flows of high order information establishing and sustaining the command and control functions of world cities.

Information flows thus summarize many other interactions, and from that perspective, it is not surprizing that recent research has attempted to redraw the boundaries of regions by looking at the flows of communications between people, for instance through mobile electronic devices (Ratti et al., 2010). Because information is tightly connected to other flows and relations, looking at its patterns allows to study the system itself.

2.2. Analysing information flows

Information flows can be studied in many ways. It is possible to study the message itself (a news item, a phone call, a letter, etc.), or the infrastructure that is necessary to carry the message (postal roads, the internet backbone, etc.).

In the case of studies focusing on the messages, newspapers have been among the first sources of data to analyse relations between cities. They were initially used to assess which distant cities were the most salient in the pages of a given city newspapers (Zipf, 1946a), and how much time it took for the information to travel between two places (Pred, 1973, 1977). This can be explained by the fact that newspapers used to be the main sources of information before the democratization of the radio and the television. In more recent periods, newspapers are in competition with many other sources of information (the radio, the television and now social media) but researchers have highlighted that they remain interesting sources. In a paper discussing the scarcity of data to study relations between cities as the global scale, Beaverstock, Smith, Taylor, Walker, and Lorimer (2000) investigate the potential of using business news taken from a city's newspapers and present this data source "as the answer to the empirical problem of studying medium-term trends in world city relations."

Until very recently, the content was still gathered manually, limiting the scope of such analysis. But an emerging trend of research, to which our paper contributes, explores ways of automating the analysis of cities through their mentions in media or textual materials (Chen, Yan, & Zhang, 2017; Meijers & Peris, 2018; Salvini & Fabrikant, 2015). While in our case we are focusing on local newspapers receiving information from distant cities (creating a directed relation), recent studies have also explored alternative methods such as looking at co-occurrences of places in documents. This methods allow to retrieve undirected relations by looking at the frequency of co-mentions of toponyms in a given textual corpus such as online news (Hu, Ye, & Shaw, 2017), Wikipedia articles (Salvini & Fabrikant, 2015) and webpages (Meijers & Peris, 2018).

Human communications can now also be analysed through exchanges between individuals on social media (Decuyper, Gandica, Cloquet, Thomas, & Delvenne, 2018) or via devices such as mobile phones. The use of phone calls to map relations is an old tradition in quantitative geography and urban planning (Board, Davies, & Fair, 1970; Zipf, 1946a), but the recent availability of very detailed datasets generated by mobile devices has given a new impulse to such studies of information exchange (Grauwin et al., 2017;Krings et al., 2009; Lambiotte et al., 2008).

Some researchers also use proxies in order to characterize collaboration networks along which information is potentially exchanged. For instance, the world city network literature (see Peris, Meijers, & van Ham, 2018 for a definition of such research schools in urban systems research) focuses on the corporate networks of big transnational firms. These scholars assume that the exchanges of high-order information and knowledge between offices of the same firms located in different places are the main driver of today's economy (Taylor & Derudder, 2015). Another interesting proxy to study exchanges of high-order information between cities is scientific collaborations that can be extracted bibliometric databases (Maisonobe, Eckert, Grossetti, Jégou, & Milard, 2016).

Finally, other studies showed that looking at the infrastructure supporting the information exchanges can also be an interesting, albeit indirect proxy of information diffusion. This is true for past networks such as the postal roads between cities (Bretagnolle & Franc, 2017) or contemporary ones such as the internet backbone (Choi, Barnett, & Chon, 2006).

2.3. Regularities in the way information travels

Despite the very different ways in how information is approached in previous studies, some of them have arrived at quite similar conclusions. Here, we list the main findings on information circulation between cities, which can generally be seen as regularities in the way information travels.

The most frequently mentioned feature in information diffusion processes is that the size of the place emitting information, and the distance between the origin of the information and its destination play a major role. This regularity has been initially discovered by Zipf (1946a) for news coverage, newspaper diffusion, phone calls and telegram messages. Since then, it has been observed with more contemporary datasets and notably mobile phone communications (Krings et al., 2009; Lambiotte et al., 2008), contradicting the idea that our contemporary society has been characterized by the "death of distance". The explanatory power of 'size' is due to the likelihood of more events occurring in bigger communities than in small communities. The importance of the second dimension – the distance – can be explained by the fact that information coming from different places will not be of the same importance as the value of a news item decreases proportionally with the distance to its origin. Zipf's model assumes that the amount of valuable information received in j from a given populated place i will be in proportion to M_i/D_{ij} where M_i is the population of place i and D_{ij} is the distance between i and j, a formulation very close to the well-known gravity model, of which several formulations exist for studying information diffusion (Grauwin et al., 2017; Krings et al., 2009; Lambiotte et al., 2008).

Studies have also highlighted the importance of administrative and cultural borders in hampering the flows of communications. This has been observed recently by researchers working on mobile phone calls in several countries (Grauwin et al., 2017). In their study they show that internal administrative borders largely reduce interactions and that such networks exhibit a strong nested hierarchical structure. The authors came up with a "hierarchical model" using the size of the interacting entities and a parameter changing with the probability for two persons from these locations to communicate, this probability being based on the "hierarchical distance" between the locations. Such impact of territorial borders has also been identified with cultural borders in countries, such as the linguistic one (Lambiotte et al., 2008), echoing other researches on the "territorial effect" (Grasland, 2010).

While working on very different datasets, researchers have highlighted very strong regularities in the way information travels. However, these studies were mostly cross-sectional and do not look at how these factors have evolved through time. In this research, we have assembled a massive dataset on interurban information flows based on recently digitalised historical local newspapers. We want to test the findings mentioned above with our dataset that covers a different geographical and temporal context, and add a cross-temporal dimension by looking at how these factors influencing the circulation of information have changed over time. Our hypotheses are:

- Hypothesis 1 (H1): Larger cities tend to be covered relatively more in news than smaller cities.
- Hypothesis 2 (H2): Geographical distance plays a role in hampering the probability of receiving news from a distant place.
- Hypothesis 3 (H3): Over time, information flows cover larger distances and the 'information field' of people increases.
- Hypothesis 4 (H4): Cultural and administrative borders hamper the circulation of information.
- Hypothesis 5 (H5): The hampering effect of cultural and administrative borders on the circulation of information has decreased over time.

3. Empirical data

3.1. Newspaper data

For the period we are studying, newspapers can be considered as the backbone of information diffusion; they were the means through which knowledge on more distant places beyond the self-experienced space was spread. For access, one did not have to be subscribed as newspapers were also hung on displays throughout the city. Newspapers were central in shaping geographical knowledge and imaginaries of the wider public (Frémont, 1976) and also played an important role in spreading relevant economic information through advertisements and price reports (Pred, 1973, 1977). As historians have argued: "Mapping the increased circulation of newspapers and mail provides a context for studying news items, adverts, tool catalogues, posters, railway timetables, and letters from loved ones – all communications from afar – and teasing out clues therein about changes in villagers' imagined geography of distant places (capital cities, the seaside, the nearest town, and sites of job prospects a long way from home)" (Schwartz, Gregory, & Thévenin, 2011).

The data on cities mentioned in historical local newspapers comes from the DIGGER dataset, a dataset that we created by using Delpher, the digital archive of historical newspapers of the Koninklijke

Bibliotheek (the National Library of the Netherlands). This dataset was built by querying massively the catalogue of the library through a search and retrieve protocol and by running named entity recognition (NER) algorithms when necessary in order to correctly identify news items containing place names. The creation of this dataset is described extensively in (Peris, Faber, Meijers, & Ham, 2020). Initially, this digital archive contained 12 million of newspapers pages. However, not all sources (newspapers) were of the same importance, for instance as they existed only for a short period. In order to build a relevant corpus we used two criteria related to the spatial and temporal coverage:

- The newspapers had to target a spatially bounded readership (this was necessary to construct origin-destination matrices);
- The newspapers had to be published in at least two consecutive decades (as we are interested in evolution of patterns of information flows)

The first criterion means that we basically excluded national newspapers. Finally, we manually removed two newspapers because many years were missing. This research focuses on the period between 1869 and the end of the year 1930. The initial time mark of this research was selected because it was the year of the abolition of a tax on newspapers - the '*dagbladzegel*' - that reduced significantly their price. The final year (1930) was chosen as it was the last census year available before the Second World War. The Second World War meant a big shock to the Dutch media landscape. Indeed, several newspapers that had been controlled by, and supported the German occupant were dissolved after the war, their place being taken by resistant journals that had started underground. Also, after the war, the share of households having a radio or a television grew very rapidly, competing with the printed press as being the main source of information.

In total, our corpus contains 31 newspapers, located in 24 cities.[1] The newspapers and their main characteristics can be found in Appendix 1. Fig. 1 shows the location of the places where they were published as well as the 312 cities and towns for which data was collected. These 312 cities and towns are the '*woonplaatsen*' that have a population of more than 10,000 inhabitants in 2013 (this obviously excludes cities in the reclaimed polders of Flevoland province that were built after our study period. More details about the selection of the settlements can be found in (Peris et al., 2020).

In this dataset, the information flows are defined as follows: First, we have a set of cities c_i, c_j, c_k, …, c_n, located in a territory. An information flow f_{ij} is occurring when the local readership of the newspaper published in the city c_j can access a news item about another city c_i. Several mechanisms can lead to such an information flow. It can be because an event occurs in c_i and the editorial board of the newspaper publish in c_j thinks that this event is of interest for the local readership, or because firms located in c_i are paying to advertise in the newspaper of c_j. The total amount of information received in city c_j from c_i for a given period t is:

$$F_{ijt} = \sum_{i \neq j}^{n} f_{ijt}$$

To give an example, Fig. 2 shows the information field of 6 different cities for the ten year period 1881–1890. The figure shows that an important majority of the information flows are spatially restricted. In the case of Amsterdam and the Hague, a great part of the flows come the main cities of North and South-Holland (Amsterdam, Rotterdam, The Hague and Utrecht), as well as smaller cities such as Haarlem, Leiden

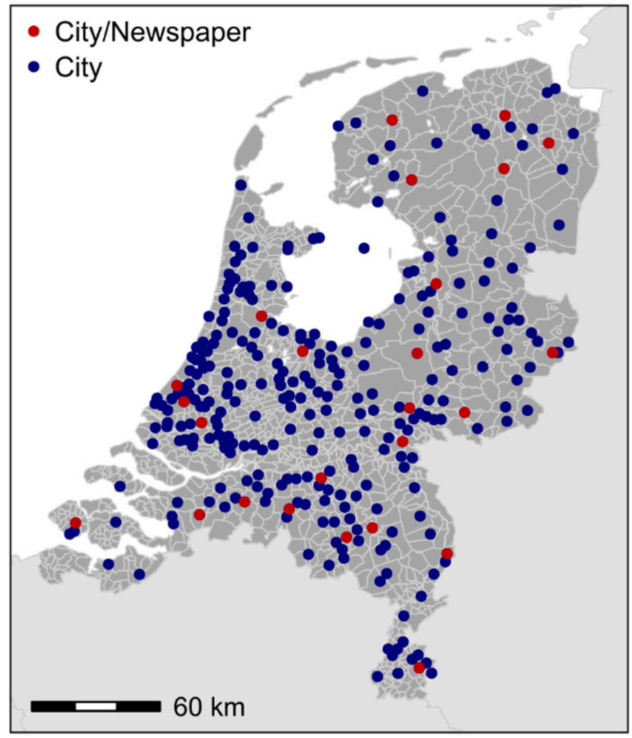

Fig. 1. Place of publication of the newspapers and cities for which the data has been collected.

and Delft. Some attention is also paid to Groningen in the North and Arnhem in the East. In the case of the newspapers from Leeuwarden, Tilburg, Venlo and Zwolle, a general patterns emerge: an important focus on the medium and small cities from their respective provinces as well a good coverage of Amsterdam. Rotterdam is also well covered, and The Hague to a lesser extent.

3.2. Historical urban populations and boundaries

Defining cities and urban population is an important step in any study on urban systems. The set of cities for which data was collected was defined according to the current population of the country. All the settlements that are above 10,000 inhabitants (a threshold commonly used by statistical agencies and researchers to define an 'urban' place) were taken. The best longitudinal data we could access is the Dutch national census which started in 1795. It is accessible for some years as linked open data through the CEDAR API[2] (Ashkpour, Meroño-Peñuela, & Mandemakers, 2015). For some missing entries or incoherent patterns (i. e. very sharp increase or decrease of the population between two years of census) we went back to the digitalized versions of the original census books[3] to manually correct the database if needed. The geometries of the municipalities were accessed through the NLGIS API.[4] As we are dealing with longitudinal data, it is important to harmonize the dataset. There are several ways for the harmonization of urban population over time depending on the initial definition of cities (morphological, political or functional), and the spatial resolution of the dataset (Cottineau, 2014). In our case, as we are dealing with irregular areal data with changing boundaries, we decided to go for a method that corrects for the municipalities that have been absorbed over the years by bigger ones. This method implies three steps:

[1] Amsterdam, Apeldoorn, Arnhem, Assen, Breda, Delft, Den Bosch, The Hague, Doetinchem, Eindhoven, Enschede, Groningen, Heerenveen, Heerlen, Helmond, Hilversum, Leeuwarden, Middelburg, Nijmegen, Roosendaal, Rotterdam, Tilburg, Venlo, Zwolle

[2] http://lod.cedar-project.nl/cedar-mini/sparql

[3] http://www.volkstellingen.nl/nl/index.html

[4] http://nlgis.nl/api/

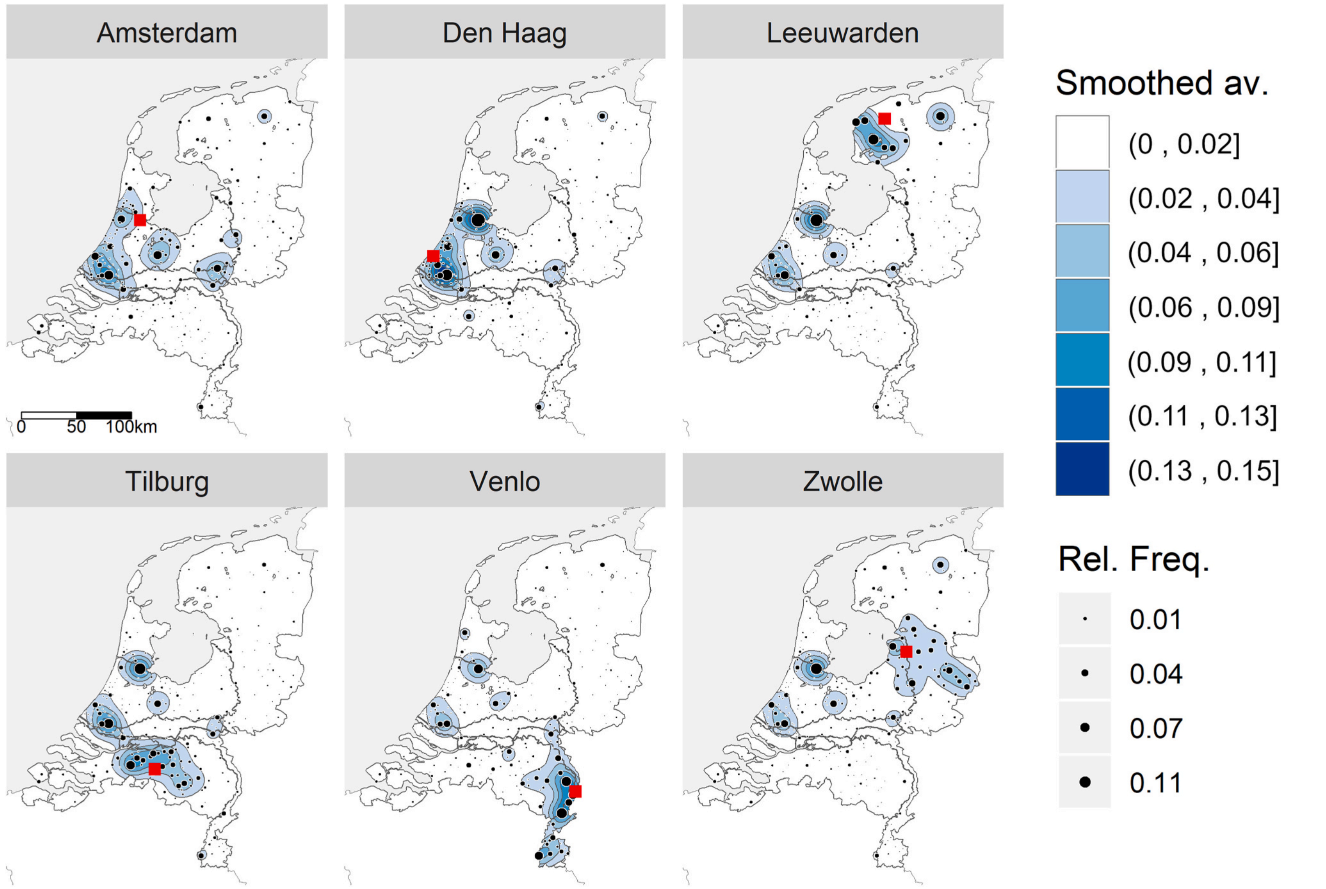

Fig. 2. Information field of 6 cities for the period 1881–1890. The red square represents the city in which the newspaper is published. To highlight clusters of cities frequently mentioned and regional focus, we computed the smoothed average with the Stewart potential function of the R package SpatialPosition (span: 10 km, function: exponential, beta: 2) (Giraud, Commenges, & Boulier, 2019). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

- The spatial delineation of cities for a baseline year (usually the last available)
- The identification of past municipalities that have been absorbed to form the city at the baseline year.
- The spatial aggregation of these municipalities under a certain threshold of population.

It is possible that two settlements are located in the same municipality. In such a case, we summed their frequency of mention in the newspapers and took the population of the entire municipality. Finally, in order to have an approximation of yearly population data, we did a linear interpolation between the different census years. This data on historical urban populations is available on a github repository.[5]

4. Research approach

4.1. A gravity framework to model changes in information diffusion

In order to systematize the analysis on such rich and multidimensional data, it is necessary to adopt a modelling framework. For that we use the gravity model, which is widespread in the studies of information flows and spatial interaction in general. The basic assumption behind this model is that flows between spatial units are proportional to the product of their size and inversely proportional to the distance between them. The origin of this model can be traced back to the middle of the 19th century, when people studying migrations and markets suggested an analogy with Newton's law of gravitation to explain the intensity of flows in the proximity of big population centres. But it is really from the 1940s that spatial interaction models of the gravity type started to spread among geographers, economists and engineers. It was for instance in contexts as different as explaining the catchment areas of universities (Stewart, 1942), intercity trips of persons (Zipf, 1946b) or information from distant places whether they come from news articles or telephone calls (Zipf, 1946a). In the following decades, the model also started to be used by engineers and planners in land use and traffic forecasting. The most common formulation of the model is:

$$F_{ij} = k\frac{M_i^{\beta} M_j^{\gamma}}{D_{ij}^{\alpha}}$$

where F_{ij} is the intensity of the interaction between i and j, M_i and M_j correspond respectively to the mass term (i.e. population, number of jobs, etc.) associated with the spatial units i and j, D_{ij} is a measure of distance between them, k is a constant of proportionality and α, β and γ are parameters to be estimated. β and γ can be understood respectively as the potential to generate and attract flows, and α is the parameter corresponding to the 'friction' of distance. The higher this exponent the more distance will play a role in lowering the intensity of interactions. While this model it nothing new, it has recently experienced a regain in interest due to the increasing availability of interaction and flow data (Grasland, 2019; Krings et al., 2009).

The gravity model also received some critiques due to its lack of universality. In a paper, Simini, González, Maritan, and Barabási (2012)

[5] https://github.com/AntoinePrs/information_field/tree/master/data

have criticized the use of adjustable parameters "that vary from region to region" and proposed a universal model – the radiation model – that is parameter-free. According to them, the radiation model could predict a wide range of interactions such as commuting, long-term migration, phone calls and freight flows with a better accuracy than the gravity model. In addition to the fact that other model comparisons did not find that the radiation model outperforms the gravity model in prediction power (Commenges, 2016; Masucci, Serras, Johansson, & Batty, 2013), the context-free dimension of the model can be seen as a limit. Indeed, according to Commenges (2016), the calibration phase in gravity modelling that Simini et al. (2012) present as a weakness, allows in fact to create knowledge on the type of spatial interaction that is studied in a given context and to incorporate this knowledge in the model. This argument is in line with the conception of this model by Burger, Oort, and Meijers (2019) which present it as a tool to "gauge" the level of interaction between spatial units. For these reasons, in order to address our hypotheses, we propose to study systematically the changes in parameters value of the gravity model calibrated with our dataset.

5. Results

5.1. Effects of size, distance and borders over the entire period

We first designed general models taking into account the entire period (1869–1930). The first model is a simple gravity type equation only taking into account the mass of the emitting city:

$$log(F_{ijt}) = k + \alpha\, log(M_{it}) + \beta\, log(D_{ij}) + \mu_j$$

Where M_i corresponds to the population of the emitting city, μ_j represents a newspaper specific fixed-effect, k is a constant and D_{ij} stand for the distance between the emitting and the receiving cities. For this model and the following ones, we use Euclidian distance. Of course, more complex ways of measuring distance could be worthwhile to consider, especially travel time between cities. However, in this paper, we are interested in the distortion from a simple isotropic situation by analysing the patterns of news flows. Further investigations on the role of infrastructure development and technological innovations on the changing patterns of these interactions would be valuable – because they would touch upon the very vivid debate on the structuring effect of transportation (Offner, 1993; Raimbault, 2020) – but they are out of the scope of this paper. For the implementation of the model we are using a multiple regression method based on an Ordinary Least Square (OLS) regression. As we are using newspapers fixed-effects, our model corresponds to what Wilson (1971) defines as an "attraction constrained" model in his family of spatial interaction models. The hypothesis behind this choice is that news can originate from cities around the publication place in a gravity model way, but the total number of news is known.

The second model introduces a variable capturing the territorial effect associated with provincial borders.

$$P_{ij} = \begin{cases} 1, & if\ p_i = p_j \\ 0, & otherwise \end{cases}$$

Where p_i is the province of city i and p_j is the province of city j. Provinces in the Netherlands are not only administrative entities. For most of them, their origin can be traced back to the medieval times. They are inherited structures with a certain degree of functional and cultural coherence. The resulting model is the following:

$$log(F_{ijt}) = k + \alpha\, log(M_{it}) + \beta log(D_{ij}) + \gamma\, P_{ij} + \mu_j$$

Finally, in the third generic model, we add a variable capturing another type of territorial effect associated with an assumed hampering role of the Rhine river. $R_{ij} = 1$ for North-North or South-South interactions, and $R_{ij} = 0$ otherwise. This variable is not only supposed to represent the physical discontinuity but also some historically grown cultural differences. For instance, it formed the northern boundary of the Roman empire, and is a gross dividing line between a majority of Catholics to the south of the river, and a majority of Protestants to the north. Being from 'below the river(s)' and 'above the rivers' is a widespread geographical reference used in daily language. The southern provinces of Limburg and North-Brabant were even incorporated later in the country. A the third model aims to catch the effect of this cultural dividing line:

$$log(F_{ijt}) = k + \alpha\, log(M_{it}) + \beta log(D_{ij}) + \gamma\, P_{ij} + \delta\, R_{ij} + \mu_j$$

Results of these three models can be found in Table 1. They show that only size and distance explain already well how news circulates. The two parameters associated with them are highly significant and 54% of the variance in the dependent variable is predicted by these two variables. As expected, city size plays a positive role on the probability of emitting news ($\alpha = 1.09$) and the distance between the news source and where the news is published plays a negative role ($\beta = -0.95$): it is less likely that news from far away appears in a local newspaper. The next two models show that being located in the same province, as well as being located on the same side of the Rhine-Meuse rivers plays a positive role on the circulation of information ($\gamma = 0.5$ and $\delta = 0.33$ for model 3). However, adding these two variables does not increase explained variance substantially due to the fact that they partly capture the distance effect. Based on these results, it seems possible to affirm that the information field of the urban readership of the cities for which we have data is spatially bounded. During the period 1869–1930, urban dwellers are more likely to encounter news from nearby and big centres, which confirms our hypothesis H1 and H2. Based on this data, we can also identify some territorial effects associated with the provincial scale and some inertia of the North-South divide that have been structuring in Dutch history. The hypothesis H3 seems also valid for the period 1869–1930.

5.2. The evolution of information circulation over time

To study the evolution of the circulation of information over time, a fourth model has been implemented. The model integrates time-varying exponents for the different factors hampering the diffusion of news.

$$log(F_{ijt}) = k + \alpha\, log(M_{it}) + \beta_t log(D_{ij}) + \gamma_t\, P_{ij} + \delta_t\, R_{ij} + \mu_j$$

As the result of this model consists of many different parameter values corresponding to the different years, they could not fit in a table. To enhance readability and interpretability, they were plotted in a diagram with the x-axis presenting the year and the y-axis the

Table 1
Results of the global models.

	Dependent variable		
	Log(Fij)		
	(1)	(2)	(3)
k	−2.009***	−3.036***	−3.425***
	(0.025)	(0.028)	(0.029)
log(Mi)	1.088***	1.089***	1.087***
	(0.002)	(0.002)	(0.002)
log(Dij)	−0.947***	−0.733***	−0.688***
	(0.003)	(0.004)	(0.004)
Pij		0.609***	0.498***
		(0.008)	(0.005)
Rij			0.331***
			(0.005)
Newspaper Fixed-effect	TRUE	TRUE	TRUE
Observations	313,995	313,995	313,995
R^2	0.545	0.552	0.557
Adjusted R^2	0.544	0.552	0.557
Residual Std. Error	1.217	1.207	1.200
F Statistic	11,727.670***	11,717.750***	11,617.010***

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

corresponding value of the parameter (Fig. 3). Using time-varying exponents improved again slightly the predictive power of the fourth model ($R^2 = 0.57$). In this first specification of a dynamic gravity model, we observe that the parameter associated with distance β_t is becoming less negative between 1869 and 1930. This would suggest a reduction of the friction of distance on the circulation of news over the time period. However, we do not observe a continuous decrease but a slightly fluctuating trend with some drops in the years before the beginning of the 20th century and at the end of the 1910s. In the case of the parameter γ_t and δ_t, associated respectively with being or not in the same province and being on the same side of the Rhine-Meuse river, we observe much more fluctuations over time. Despite these important fluctuations, we can observe a slightly increasing tendency for both γ_t and δ_t. Such results are paradoxical because while β_t is becoming less negative, indicating an increasing probability of receiving news from far away cities, γ_t and δ_t, that are associated with short distance interactions, are increasing. One possible interpretation of these changes in parameters values would be that both short and long distances interactions are increasing, while middle range distance interactions are decreasing. The intermediate places would be sort of jumped over. This interpretation would be in line with previous works on "space-time contraction" that show that in the case of transportations networks, the increase in speed is associated with fewer stops and a weakening of intermediate places and smaller cities (Bretagnolle & Pumain, 2010; Janelle, 1968).

Finally, we implemented a fifth model with a time-varying parameter associated with the mass of the cities M_{it}. This parameter corresponds to the potential of generating information in respect to size.

$$log\left(F_{ijt}\right) = k + \alpha_t\ log(M_{it}) + \beta_t log\left(D_{ij}\right) + \gamma_t\ P_{ij} + \delta_t\ R_{ij} + \mu_j$$

The goodness of fit of this model is similar to the previous one with $R^2 = 0.57$. Surprisingly, the evolution of the parameter associated with distance is opposite. This means that when controlling for the changing potential to generates news relative to the size, the general tendency indicates a growing probability of short distance interactions. This is not necessarily contradictory with the previous interpretation as part of the long distance interactions that are associated with big cities and its impact on the β_t could be captured by the parameter α_t.

However, we can observe that the trend in the evolution of parameters in a longitudinal gravity model is largely influenced by the specification of the equation and the variables taken into account. This has already been raised by the economic literature that focuses on international trade and has been coined the "distance puzzle" (Brun, 2005). Further works on the applicability of the gravity model in cross-temporal analysis are needed. This also shows the limits of focusing on a single indicator such as the distance friction to summarize complex spatial phenomena. No definitive conclusions can be derived from these dynamic gravity models. However, they provide some hints on a possible space-time contraction and that effects found are not general for all cities, instead suggesting that there is spatial heterogeneity that is hard to capture with general models.

6. Spatial heterogeneity and city trajectories

We have seen in the preceding section that the gravity model does not allow looking at dynamics of individual cities, even though some results suggested that effects may differ for different types of cities. In the following subsections we will look at disaggregated data at the city scale in order to better understand some dynamics suggested by the models. We will first look at the average distance travelled by information flows in the form of news items.

6.1. Average distance of information flows

An alternative way of looking at the changes in the spatial extent of the information field of cities is to compute the average distance of information flows for every year and every city for which we have one or more newspapers. For any city j, the index will be computed the following way:

$$\overline{D}_{jt} = \frac{1}{n}\sum_{i=1}^{n} f_{ijt}.d_{ij}$$

For example, if a newspaper based in Delft publishes two news items about Rotterdam (located at 12.9 km), one about The Hague (8.3 km) and one about Amsterdam (54.4 km) at time t, $\overline{D}_{jt} = 22.125$. The results of this index are shown in Fig. 4. In order to compare the changes for different cities, we have subtracted the value of the first year to all other years. This way, the lines depart from 0. The value of the initial year D (t1) is presented also in the figure. While this first index is characterized by strong disparities, they are not interesting to interpret as they are mostly related to the relative position of a city toward all other cities. Therefore Amsterdam, Rotterdam, Hilversum and Arnhem have lower values than Eindhoven, Venlo and Groningen that are located close to the borders. Of greater interest is the trend of these lines. For this, we computed a linear regression where the average distance of information flows is considered as a linear function of time. The parameter associated with the slope is displayed on Fig. 4. One of the first conclusions that can be drawn is that except for two cases (Rotterdam and Middelburg), the spatial information field of all cities have expanded over the period we are studying. This result tends to confirm our hypothesis about the fact that information flows occur over larger distances (H3).

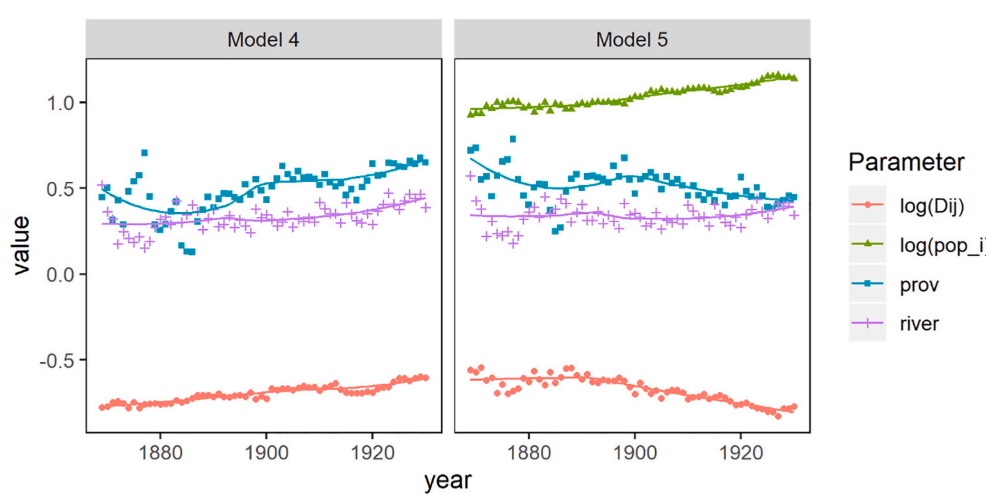

Fig. 3. Results of the models 4 and 5.

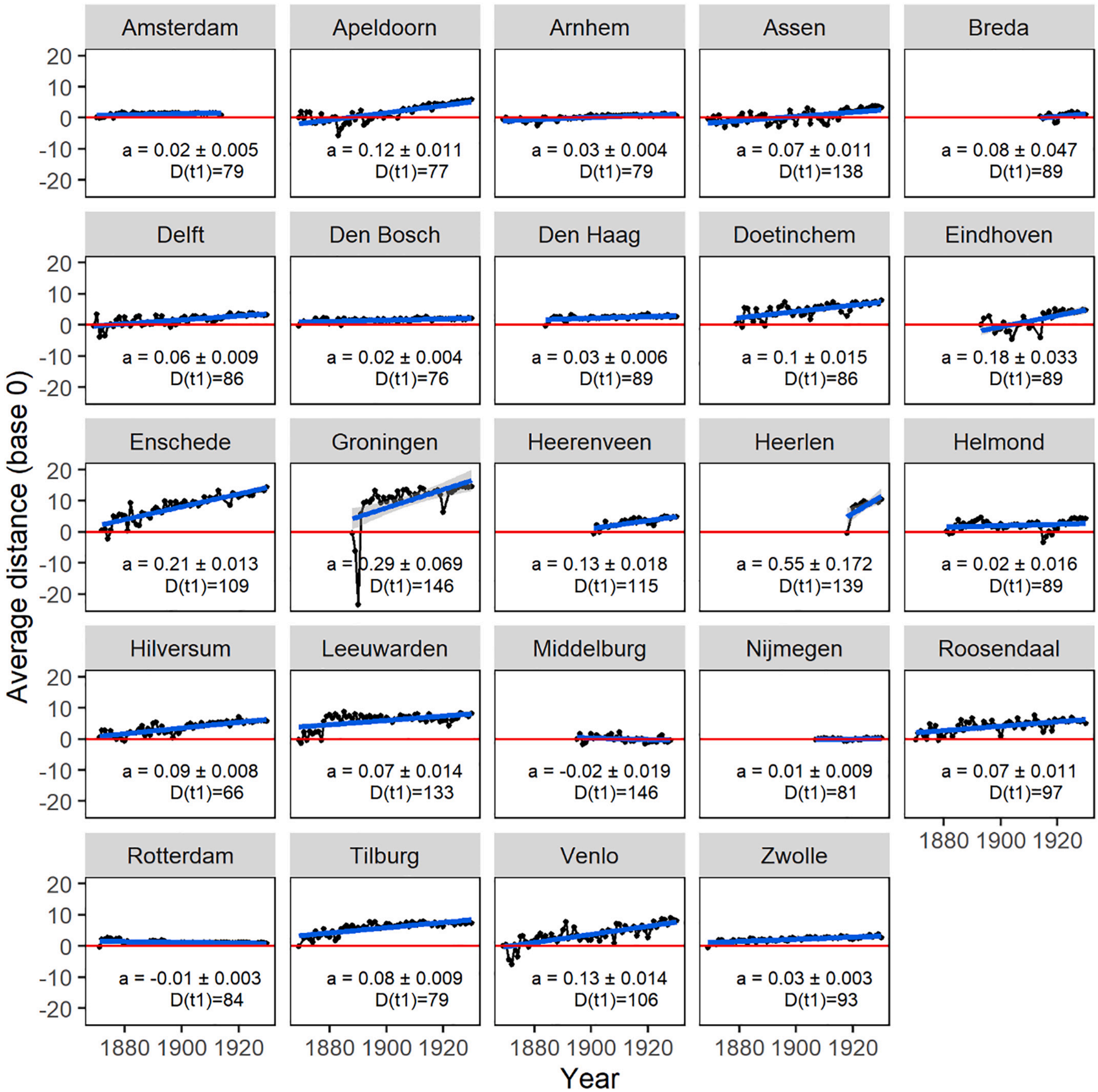

Fig. 4. Yearly evolution of the average distance of information flows for each cities.

However, there are important variations in the trend of these time series. The cities that have the most increasing trend are Heerlen (0.55), Groningen (0.29), Enschede (0.21), Eindhoven (0.18), Venlo (0.13), Heerenveen (0.13) and Doetinchem (0.1). These cities are all located on the outskirts of the country in the Northern provinces of Friesland and Groningen, in the South (Limburg and North-Brabant) or in the most western parts of the eastern province Gelderland and Overijssel. At the opposite of this trend, cities from the core area of what is today known as the Randstad (composed of the most urbanized areas of North- and South-Holland) have flatter trends. This is especially the case for the big centres such as Amsterdam (0.02), Den Haag (0.03) and Rotterdam (−0.01), but not for the smallest centres such as Delft (0.06) and Hilversum (0.09). So, in peripheral cities interest in the largest cities in the core increased, but this was hardly true vice versa. This trend corresponds to the change in the Dutch urban system described by van Engelsdorp Gastelaars and Wagenaar (1981) as the 'rise of the Randstad'. For these authors the process of political, cultural and economic integration of the Netherlands that took place at the end of the 19th century and in the beginning of the 20th century has resulted in an increasing polarization toward the core (the Randstad) and a rising contrast between this region and the peripheries (the eastern, northern and southern regions). This process can be related to what van der van der Knaap (1980) has observed in his study of the Dutch urban system between 1840 and 1970. For him, the development of the inter-city linkages has resulted in an increasing concentration of the population in the largest cities (Amsterdam, Rotterdam, The Hague and Utrecht), where 25% of the population of the Netherlands were living in 1910, when the polarization was the highest. Our data shows well this polarization process with peripheral cities reporting increasingly more often about distant Randstad cities, while the average distance of news flows to cities in the Randstad core region does not change much.

6.2. The spatial distribution of origins of information flows

While the average distance of information flows is an interesting index to look at rather general trends, its aggregated dimension does not allow to describe the spatial complexity of the changes in information field. For instance, as Fig. 5 shows, in the case of the information field of Enschede, one can observe three different processes. First, the close-by cities of Almelo and Hengelo, as well as smaller nearby towns, remain very well covered in all three periods mapped. Second, more and more information from the big cities of the Randstad, and especially The Hague are published. The latter may have to do with the rising importance of the nation state, and The Hague is the seat of Dutch central government. Finally, the intermediate cities of Deventer and Arnhem, that used to be important sources of information in the period 1871–1890 are less important in the period 1911–1930. A quite different pattern can be observed in the case of Rotterdam. The drop of long distance information flows coming from Maastricht, Den Bosch, Nijmegen and Breda tends to confirm the pattern revealed by the average distance index. For more close by interactions, one can see that Den Haag skyrocketed as a source of information for Rotterdam readers.

To systematize this analysis for all the cities for which we have newspapers, we plotted the kernel density distribution of the distance travelled by information flows for 4 different periods of time (1870–1884, 1885–1899, 1900–1914, 1915–1929). Fig. 6 shows the changes in coverage with respect to distance. One of the most striking patterns that emerges from this visualisation is the very clear increase of information flows coming from the Randstad area (the vertical red lines indicate the distance to the four largest Dutch cities Amsterdam, Rotterdam, The Hague and Utrecht). This is true for almost all the cities except Delft, which tends to cover less and less these big urban centres and Hilversum, which focuses less on nearby Amsterdam and Utrecht but more on Rotterdam and The Hague. Most of the other cities for which we have newspapers tend to focus more and more on the Randstad area. It demonstrates the rise of an economic and political core region in the Netherlands. This attention for nearby urban centres, including close-by medium-sized cities (represented by a green line in the Figure), decreases in relative terms. This is especially visible in the case of Apeldoorn, Arnhem, Den Bosch, Doetinchem, Eindhoven, Enschede, Leeuwarden, Tilburg and Venlo and Roosendaal where the density polygons of the most recent years are very clearly more and more inflated toward the right end of the x-axis of the graph. Notable exceptions to this tendency are Helmond and Groningen that receive more news flows from nearby cities and towns.

When analysing the pattern for the three Randstad cities we have data for, an opposite trend can be observed. For the three of them, the cities in the immediate proximity receive more and more attention. This might be caused by what van Engelsdorp Gastelaars and Wagenaar (1981) refer to as the “suburbanisation of the Randstad-centres” that happened at the end of the period we are studying. This phenomenon is described as the expansion of the daily system of the big urban centres due to the congestion in the core city generated by sharp increase of economic activities and population. This could also explain the stagnation or decrease of the average distance of news flows observed previously for the big Randstad cities, with the small settlements located in the periphery of big centres receiving more attention at the same time as being incorporated in wider functional entities.

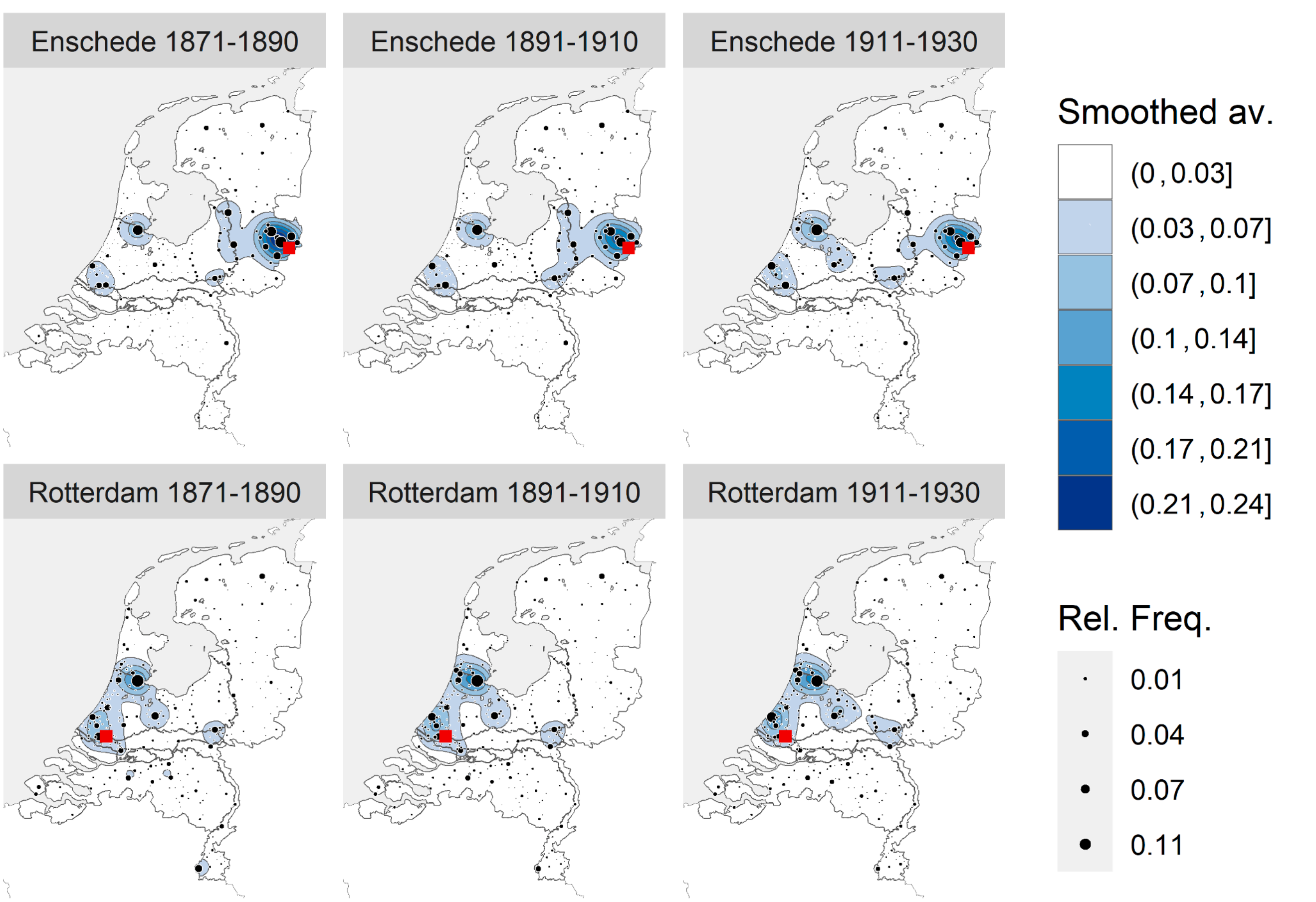

Fig. 5. Evolution of the information field of Enschede and Rotterdam. Smoothed averages were computed with the Stewart potential function of the R package SpatialPosition (span: 11 km, function: exponential, beta: 2) (Giraud et al., 2019).

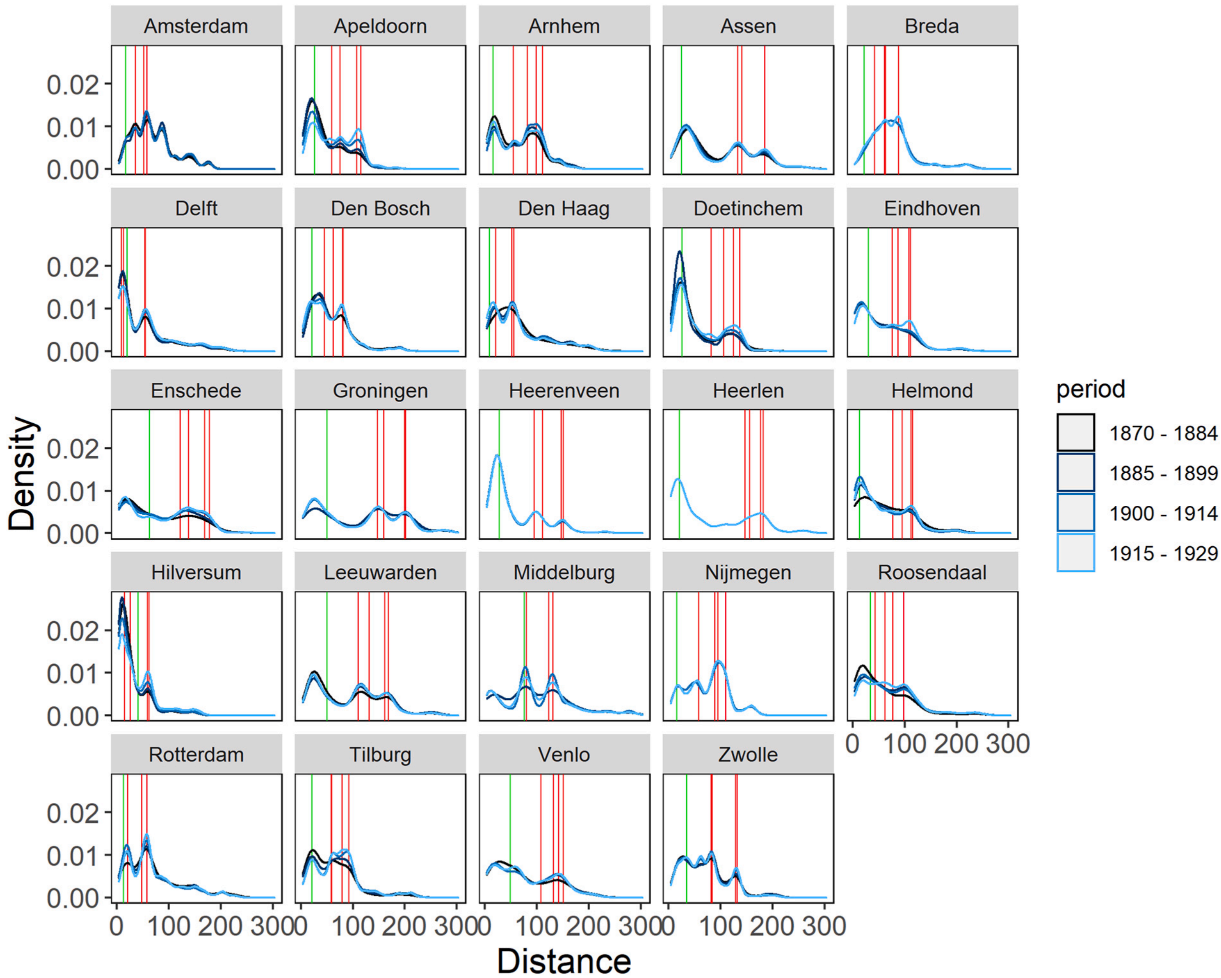

Fig. 6. Kernel density distribution of the origin of information flows. The vertical red lines represent the location of the four big cities of the Randstad (Amsterdam, Rotterdam, The Hague and Utrecht). The green line represents the closest important city (in the top 15 most populated place at least once during the period), that is not located in the Randstad. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

7. Conclusions and discussions

Our findings tend to confirm previous research that shows that space matters greatly in the process of diffusion of information. We confirmed that for the period between the 1869 and 1930, the size of cities emitting information and the distance between cities are important and significant factors explaining the circulation of information (as hypothesised in H1 and H2). Moreover, the hampering dimension of provincial borders and the North-South divide that have been structuring the Netherlands for a long time remained important over the entire period (as hypothesised by H4).

While we could achieve relatively good predictions with the dynamic gravity models used to analyse the evolution of the circulation of information, we could not draw clear conclusions on the influence of the different factors over time. At this stage of the analysis, we could not confirm the hypotheses H3 and H5 about the evolution of the hampering effect of distance and borders. However, more descriptive analysis on the average distance over which information travels as well as the spatial distribution of the origins of information flows allowed us to confirm the existence of a space-time contraction at this period (H3). We could also identify one of the main driving factors behind this increase of long distance interactions. Our analysis of the information flows between the Dutch cities revealed that the period is clearly characterized by a polarization around its main economic, political and demographic core that from 1938 onwards would be referred to as the "Randstad" (Meijers, 2019), and that became a major focus of the Dutch planning debate in the following decades (van Meeteren, 2020). Newspapers published outside of this area tend to report increasingly often on the Randstad cities, most of the time at the expense of the closer-by medium-sized cities. This process can be related to the hierarchical selection within the urban system in the context of space-time contraction (Bretagnolle & Pumain, 2010; Janelle, 1968). Such a process has also been observed by van der van der Knaap (1980) in a study on the Dutch urban system that also relates changes in the urban hierarchy to the "increasing scale of the spatial organisation of society".

The data allows for much more analysis and is available for other researchers (Peris et al., 2020). It could for instance be worthwhile to make a clear link with media studies, for instance through exploring the association between the political/religious orientation of newspapers and their spatial information field. Also, while we limited our analysis to simple counts of news items, it may be valuable to also incorporate the content of these news items so that information flows could be classified. This would require employing more advanced content analysis techniques and machine learning. We see a particularly interesting challenge in relating the image (positive/negative) that is created of particular cities elsewhere and how this has influenced the spatial behaviour of firms and individuals, for instance with respect to migration. Further research could also explore how changes in transportation technology influence the patterns of information diffusion. Finally, the pattern of co-occurrences of place names in news items could be used to identify the relationships between cities, in a similar vein as was recently done for the CommonCrawl Web Archive (Meijers & Peris, 2018).

To our knowledge, this research is among the first systematic explorations of the geographic dimension of digitalised, historical text archives, and certainly the first to explore the spatial dimension of the digitalised newspaper archive in the Netherlands, a resource that has so far been the exclusive domain of the digital humanities. What we essentially show is the feasibility of using a computational social science approach to construct completely novel geographically relevant data sets, allowing us to reconstruct the (evolution of) the spatial organisation of a territory over time. Throughout the world, many programs are currently investing in the digitalisation of such archives. This means that our experiences could be relevant for those wanting to exploit the wealth of geographical information hidden in them.

Declaration of Competing Interest

The authors declare no conict of interest.

Acknowledgement

This work was funded through a VIDI grant (452-14-004) provided by the Netherlands Organisation for Scientific Research (NWO), and through the researcher-in-residence program of the Koninklijke Bibliotheek (KB), the national library of the Netherlands. The authors would like to thank the staff of the research department of the KB for their support and welcome.

Appendix 1. List of newspapers included in the study

ID	Title	Publication place	First year	Last year	News items
37631091X	De maasbode	Rotterdam	1871	1939	1,110,870
376311770	Tilburgsche cr.	Tilburg	1869	1931	668,214
376312912	Delftsche cr.	Delft	1866	1945	520,290
398540756	Pr. Drentsche en Asser cr.	Assen	< 1865	1950	566,305
398541485	Haagsche courant	Den Haag	1884	1939	1,294,979
398543062	Pr. Overijsselsche en Zwolsche cr.	Zwolle	< 1865	1944	867,612
398825769	De grondwet	Roosendaal	1870	1955	335,136
398831475	Pr. Noordbrabantsche en 's Hertogenbossche cr.	Den Bosch	< 1865	1941	866,093
398831920	Tubantia	Enschede	1872	1942	400,711
399290591	Nieuwe Apeldoornsche cr.	Apeldoorn	1911	1945	250,668
400336960	Pr. Geldersche en Nijmeegsche cr.	Nijmegen	1907	1941	392,495
400337010	Venloosch weekblad	Venlo	< 1865	1898	64,013
400337029	Venloosche cr.	Venlo	1887	1908	52,521
400337088	Apeldoornsche cr.	Apeldoorn	< 1865	1924	210,547
400337266	Eindhovensch dagblad	Eindhoven	1914	1938	215,873
400337274	De Peel- en Kempenbode	Eindhoven	1893	1911	61,536
400337282	De Zuid-Willemsvaart	Helmond	1881	1944	294,252
400337452	Bredasche cr.	Breda	1914	1939	178,837
400337789	Arnhemsche cr.	Arnhem	< 1865	1950	783,105
400383756	Nieuwe Tilburgsche Cr.	Tilburg	1879	1944	1,012,536
400915138	Nieuwe Venlosche cr.	Venlo	1909	1941	241,965
401028933	Limburger koerier : pr. dagblad	Heerlen	1920	1975	546,984
832005797	Nieuwsblad van Friesland	Heerenveen	1901	1994	479,167
832401439	De Gooi- en Eemlander	Hilversum	1871	1950	554,704
83245351X	Limburgsch dagblad	Heerlen	1918	1994	470,076
83249562X	Het nieuws van den dag	Amsterdam	1870	1914	1,697,465
832564818	Rotterdamsch nieuwsblad	Rotterdam	1878	1944	2397235
832915580	De Graafschap-bode	Doetinchem	1879	1947	583,000
833013246	Nieuwsblad van het Noorden	Groningen	1888	1994	1,298,678
852115210	Leeuwarder courant	Leeuwarden	< 1865	1994	1,457,970
852121741	Middelburgsche courant	Middelburg	< 1865	1928	470,537

References

Ashkpour, A., Meroño-Peñuela, A., & Mandemakers, K. (2015). The aggregate Dutch historical censuses. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 48*(4), 230–245. https://doi.org/10.1080/01615440.2015.1026009.

Batty, M. (2013). *The new science of cities*. MIT Press.

Beaverstock, J. V., Smith, R. G., Taylor, P. J., Walker, D. R. F., & Lorimer, H. (2000). Globalization and world cities: Some measurement methodologies. *Applied Geography, 20*(1), 43–63. https://doi.org/10.1016/S0143-6228(99)00016-8.

Board, C., Davies, R. J., & Fair, T. J. D. (1970). The structure of the South African space economy: An integrated approach. *Regional Studies, 4*(3), 367–392. https://doi.org/10.1080/09595237000185361.

Bod, R. (2013). Who's afraid of patterns?: The particular versus the universal and the meaning of humanities 3.0. *BMGN-Low Countries Historical Review, 128*(4), 171–180.

Bretagnolle, A., & Franc, A. (2017). Emergence of an integrated city-system in France (XVIIth–XIXth centuries): Evidence from toolset in graph theory. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 50*(1), 49–65. https://doi.org/10.1080/01615440.2016.1237915.

Bretagnolle, A., & Pumain, D. (2010). Simulating urban networks through multiscalar space-time dynamics: Europe and the United States, 17th-20th centuries. *Urban Studies, 47*(13), 2819–2839. https://doi.org/10.1177/0042098010377366.

Brun, J.-F. (2005). Has distance died? Evidence from a panel gravity model. *The World Bank Economic Review, 19*(1), 99–120. https://doi.org/10.1093/wber/lhi004.

Burger, M., van Oort, F., & Meijers, E. (2019). Examining spatial structure using gravity models. In L. D'Acci (Ed.), *The mathematics of urban morphology (pp. 471–479)*. Springer International Publishing. https://doi.org/10.1007/978-3-030-12381-9_21.

Castells, M. (1996). *The rise of the network society: The information age: Economy, society and culture*. Wiley.

Chen, Y., Yan, F., & Zhang, Y. (2017). Local name, global fame: The international visibility of Chinese cities in modern times. *Urban Studies, 54*(11), 2652–2668. https://doi.org/10.1177/0042098016646674.

Choi, J. H., Barnett, G. A., & Chon, B.-S. (2006). Comparing world city networks: A network analysis of internet backbone and air transport intercity linkages. *Global Networks, 6*(1), 81–99. https://doi.org/10.1111/j.1471-0374.2006.00134.x.

Commenges, H. (2016). Modèle de radiation et modèle gravitaire—Du formalisme à l'usage. *Revue Internationale de Géomatique, 26*(1), 79–95. https://doi.org/10.3166/RIG.26.79-95.

Cottineau, C. (2014). *L'évolution des villes dans l'espace post-soviétique. Observation et modélisations.* [Phdthesis, Université Paris 1 Panthéon-Sorbonne]. https://tel.archives-ouvertes.fr/tel-01167818/document.

Decuyper, A., Gandica, Y., Cloquet, C., Thomas, I., & Delvenne, J.-C. (2018). Measuring the effect of node aggregation on community detection. *ArXiv:1809.08855 [Physics]*. http://arxiv.org/abs/1809.08855.

Derudder, T., Witlox, & Catalano. (2003). Hierarchical Tendencies and Regional Patterns in the World City Network: A Global Urban Analysis of 234 Cities. *Regional Studies, 37*(9), 875–886. https://doi.org/10.1080/0034340032000143887.

van Engelsdorp Gastelaars, R., & Wagenaar, M. (1981). *The rise of the 'Randstad', 1815–1930. In Patterns of European Urbanisation Since 1500. Routledge*.

Frémont, A. (1976). *La région, espace vécu*. Presses universitaires de France.

Giraud, T., Commenges, H., & Boulier, J. (2019). *SpatialPosition: Spatial Position Models* (2.0.0) [Computer software]. https://CRAN.R-project.org/package=SpatialPosition.

Grasland, C. (2010). Spatial Analysis of Social Facts. In Handbook of Quantitative and Theoretical Geography. Advances in Quantitative and Theoretical Geography (pp. 000–046). Faculty of the Geosciences and Environment of the University of Lausanne. https://halshs.archives-ouvertes.fr/halshs-00410669.

Grasland, C. (2019). *International news flow theory revisited through a space–time interaction model: Application to a sample of 320,000 international news stories published through RSS flows by 31 daily newspapers in 2015. International Communication Gazette, 1748048518825091*. https://doi.org/10.1177/1748048518825091.

Grauwin, S., Szell, M., Sobolevsky, S., Hövel, P., Simini, F., Vanhoof, M., … Ratti, C. (2017). Identifying and modeling the structural discontinuities of human interactions. *Scientific Reports, 7*, Article 46677. https://doi.org/10.1038/srep46677.

Hägerstrand, T. (1967). *Innovation diffusion as a spatial process*. University of Chicago Press.

Hohenberg, P. M., & Lees, L. H. (2009). *The making of urban Europe* (pp. 1000–1994). Harvard University Press.

Hu, Y., Ye, X., & Shaw, S.-L. (2017). Extracting and analyzing semantic relatedness between cities using news articles. *International Journal of Geographical Information Science, 31*(12), 2427–2451. https://doi.org/10.1080/13658816.2017.1367797.

Janelle, D. G. (1968). Central place development in a space-time framework. *The Professional Geographer, 20*(1), 5–10. https://doi.org/10.1111/j.0033-0124.1968.00005.x.

van der Knaap, G. A. (1980). *Population growth and urban systems development: A case study* (M. Nijhoff).

Krings, G., Calabrese, F., Ratti, C., & Blondel, V. D. (2009). Urban gravity: A model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment, 2009*(07), L07003. https://doi.org/10.1088/1742-5468/2009/07/L07003.

Lambiotte, R., Blondel, V. D., de Kerchove, C., Huens, E., Prieur, C., Smoreda, Z., & Van Dooren, P. (2008). Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications, 387*(21), 5317–5325. https://doi.org/10.1016/j.physa.2008.05.014.

Lansdall-Welfare, T., Sudhahar, S., Thompson, J., Lewis, J., Team, F. N., & Cristianini, N. (2017). Content analysis of 150 years of British periodicals. *Proceedings of the National Academy of Sciences, 114*(4), E457–E465. https://doi.org/10.1073/pnas.1606380114.

Maisonobe, M., Eckert, D., Grossetti, M., Jégou, L., & Milard, B. (2016). The world network of scientific collaborations between cities: Domestic or international dynamics? *Journal of Informetrics, 10*(4), 1025–1036. https://doi.org/10.1016/j.joi.2016.06.002.

Masucci, A. P., Serras, J., Johansson, A., & Batty, M. (2013). Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. *Physical Review E, 88*(2), Article 022812. https://doi.org/10.1103/PhysRevE.88.022812.

van Meeteren, M. (2020). A prehistory of the polycentric urban region: Excavating Dutch applied geography, 1930–60. *Regional Studies*, 1–14. https://doi.org/10.1080/00343404.2020.1800629.

Meier, R. L. (1962). *A communications theory of urban growth*. Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University.

Meijers, E., & Peris, A. (2018). Using toponym co-occurrences to measure relationships between places: Review, application and evaluation. *International Journal of Urban Sciences*, 1–23. https://doi.org/10.1080/12265934.2018.1497526.

Meijers, E. J. (2019). Herkomst van het concept Randstad: De strijd om de locatie van de nationale luchthaven. *Geografie, Januari, 2019*, 32–33.

Meijers, E. J., Burger, M. J., & Hoogerbrugge, M. M. (2016). Borrowing size in networks of cities: City size, network connectivity and metropolitan functions in Europe. *Papers in Regional Science, 95*(1), 181–198. https://doi.org/10.1111/pirs.12181.

Offner, J.-M. (1993). Les « effets structurants » du transport: Mythe politique, mystification scientifique. *L'Espace géographique, 22*(3), 233–242. https://doi.org/10.3406/spgeo.1993.3209.

Peris, A., Faber, W. J., Meijers, E., & van Ham, M. (2020). One century of information diffusion in the Netherlands derived from a massive digital archive of historical newspapers: The DIGGER dataset. *Cybergeo: European Journal of Geography*. http://journals.openedition.org/cybergeo/33747.

Peris, A., Meijers, E., & van Ham, M. (2018). The evolution of the Systems of Cities Literature since 1995: Schools of thought and their interaction. *Networks and Spatial Economics, 18*(3), 533–554. https://doi.org/10.1007/s11067-018-9410-5.

Pred, A. (1973). *Urban growth and the circulation of information: The United States system of cities* (pp. 1790–1840). Harvard University Press.

Pred, A. (1977). *City Systems in Advanced Economies: Past growth, present processes, and future development options*. Wiley.

Pumain, D. (2011). Systems of cities and levels of organisation. In P. Bourgine, & A. Lesne (Eds.), *Morphogenesis* (pp. 225–249). Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-642-13174-5_13.

Raimbault, J. (2020). Hierarchy and co-evolution processes in urban systems. *ArXiv: 2001.11989 [Physics]*. http://arxiv.org/abs/2001.11989.

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., … Strogatz, S. H. (2010). Redrawing the map of Great Britain from a network of human interactions. *PLoS One, 5*(12), e14248. https://doi.org/10.1371/journal.pone.0014248.

Salvini, M. M., & Fabrikant, S. I. (2015). Spatialization of user-generated content to uncover the multirelational world city network. *Environment and Planning B: Planning and Design.*. https://doi.org/10.1177/0265813515603868.

Sassen, S. (1991). *The Global City: New York, London, Tokyo*. Princeton University Press.

Schwartz, R., Gregory, I., & Thévenin, T. (2011). Spatial history: Railways, uneven development, and population change in France and Great Britain, 1850-1914. *The Journal of Interdisciplinary History, 42*(1), 53–88.

Simini, F., González, M. C., Maritan, A., & Barabási, A.-L. (2012). A universal model for mobility and migration patterns. *Nature, 484*(7392), 96–100. https://doi.org/10.1038/nature10856.

Stephens, M., & Poorthuis, A. (2015). Follow thy neighbor: Connecting the social and the spatial networks on twitter. *Computers, Environment and Urban Systems, 53*, 87–95. https://doi.org/10.1016/j.compenvurbsys.2014.07.002.

Stewart, J. Q. (1942). A measure of the influence of a population at a distance. *Sociometry, 5*(1), 63. https://doi.org/10.2307/2784954.

Taylor, P. J. (2001). Specification of the world city network. *Geographical Analysis, 33*(2), 181–194.

Taylor, P. J., & Derudder, B. (2015). *World City network: A global urban analysis*. Routledge.

Törnqvist, G. (1968). Flows of information and the location of economic activities. *Geografiska Annaler. Series B, Human Geography, 50*(1), 99–107. https://doi.org/10.2307/490320.

Törnqvist, G. (1970). *Contact systems and regional development. Lund*. University Press.

Wilson, A. G. (1971). A family of spatial interaction models, and associated developments. *Environment and Planning A, 3*(1), 1–32.

Zipf, G. K. (1946a). Some determinants of the circulation of information. *The American Journal of Psychology, 59*(3), 401–421. https://doi.org/10.2307/1417611.

Zipf, G. K. (1946b). The P1 P2/D Hypothesis: On the Intercity Movement of Persons. *American Sociological Review, 11*(6). https://doi.org/10.2307/2087063, 677–686. JSTOR.