Central Places in Wikipedia

Carsten Keßler

Abstract Central Place Theory explains the number and locations of cities, towns, and villages based on principles of market areas, transportation, and socio-political interactions between settlements. It assumes a hexagonal segmentation of space, where every central place is surrounded by six lower-order settlements in its range, to which it caters its goods and services. In reality, this ideal hexagonal model is often skewed based on varying population densities, locations of natural features and resources, and other factors. In this paper, we propose an approach that extracts the structure around a central place and its range from the link structure on the Web. Using a corpus of georeferenced documents from the English language edition of Wikipedia, we combine weighted links between places and semantic annotations to compute the convex hull of a central place, marking its range. We compare the results obtained to the structures predicted by Central Place Theory, demonstrating that the Web and its hyperlink structure can indeed be used to infer spatial structures in the real world. We demonstrate our approach for the four largest metropolitan areas in the United States, namely New York City, Los Angeles, Chicago, and Houston.

1 Introduction

Central Place Theory was developed in the 1930s, following the observation of recurring patterns in the arrangement of settlements of different sizes (Christaller, 1933, Baskin, 1966).[1] It explains the number and locations of cities, towns, and villages based on principles of market areas, transportation, and socio-political interactions between settlements. Under perfect – and somewhat unrealistic – conditions, Central Place Theory predicts a hexagonal segmentation of space, such that six lower-order settlements (e.g., towns) arrange around one higher-order settlement (e.g., a city). These purely spatial explanations of Central Place Theory were later extended to take economic considerations, such as competition, into account (Lösch, 1954). Sizes and market areas of the respective settlements are not fixed, but rather depend on population density, locations of natural features and resources, and other factors.

Central Place Theory has been shown to apply in a number of places, especially when the local situation is close to the underlying assumptions of the theory (Berry and Garrison,

Carsten Keßler
Center for Advanced Research of Spatial Information and Department of Geography
Hunter College, City University of New York
New York, USA
e-mail: carsten.kessler@hunter.cuny.edu

[1] The title of the current paper alludes to the title of the original publication introducing Central Place Theory by Christaller (1933), translated by Baskin (1966): *Central Places in Southern Germany*.

1958b, Brush, 1953, for example). Even in cases where the spatial arrangement of the settlements cannot be easily explained by Central Place Theory, the formation of lower-order settlements around central places that provide certain goods or services is still evident and can be observed everywhere in the developed world. As such, Central Place Theory explains networks of dependencies, where smaller settlements depend on goods, services, and the job markets of a larger settlement in their vicinity.

The premise of the research presented here is that those central places and the spatial configuration of settlements in their range can be inferred from the link structure on the Web. Using a corpus of georeferenced documents from the English language edition of Wikipedia, our results indicate that these dependencies between smaller and larger settlements are reflected in the number of references between their corresponding Wikipedia pages. The underlying assumption is that a central place will be referred to more often, specifically from places in its range that are functionally dependent.

Following this approach, we assess the range of a central place based on the frequency distribution of the distances to *referring* places, i.e., places whose corresponding Wikipedia pages link to this place, or mention it in the text. We assign weights to the incoming links from other places based on the count of references on their pages; for example, the Wikipedia page for Jersey City, New Jersey, contains one hyperlink to the page for New York City, and 9 mentions of the term “New York City” in its text, resulting in a total of 10 references from Jersey City to New York City. In comparison, the page for Hoboken, New Jersey, contains 4 hyperlinks and 49 mentions, resulting in a total of 53 references. We show how these reference counts can be employed as weights in our model to account for the relative importance (or *centrality*) of New York City for Jersey City and Hoboken, respectively.

Like most other studies that employ the Web as a source for geographic information, we face the Geoweb Scale Problem (Hecht and Moxley, 2009): The geometry of each place is only available as a point coordinate in Wikipedia, independent of whether the represented feature is as small as a statue in Central Park, or as large as New York State. Therefore, the semantics of the relationships between two places plays an important role, in addition to using the number of references for the weighting. We take into account the administrative hierarchy between places obtained from the GeoNames gazetteer[2] and the DBpedia ontology (Lehmann et al, 2012) as a filter to tackle this problem. We demonstrate the general feasibility of this combined approach by analyzing the link structure of the four largest regional capitals in the United States, namely New York City, Los Angeles, Chicago, and Houston.

The remainder of this paper is organized as follows: In the next section, we discuss relevant related work on Central Place Theory and on the analysis of Wikipedia contents. Section 3 describes the process of obtaining and preparing the dataset used in this study, followed by a specification of the main characteristics of the dataset in Section 4. The process of identifying central places in our dataset is introduced in Section 5, including a detailed discussion of the different choices made in the process. Section 6 discusses the obtained results, followed by concluding remarks in Section 7.

2 Related Work

This section reviews the core ideas of Central Place Theory and gives an overview of relevant related work on the analysis of geospatial content on the Web, with a focus on Wikipedia.

[2] `http://www.geonames.org`

2.1 Central Place Theory

Central Place Theory (Christaller, 1933, Baskin, 1966) explains the spatial configurations of central places from a purely economic perspective, viewing them mainly as locations where people come together to trade goods and services. The theory is based on a number of assumptions, such as an isotropic plane, evenly distributed population and resources, profit-oriented sellers, and economic customers who aim to minimize their travel to obtain goods. Lösch (1954) later relaxed this rigid economic perspective, modifying the theory to optimize for consumer welfare. The rank order of central places distinguishes hamlets (first-order centers), villages (second-order centers), towns (third-order centers), cities (fourth-order centers), and regional capitals (fifth-order centers). Figure 1 shows the hexagonal spatial configuration arising from these five orders of centers. According to this categorization, all four places that we investigate here are regional capitals, and we try to identify the depending cities and towns around them.

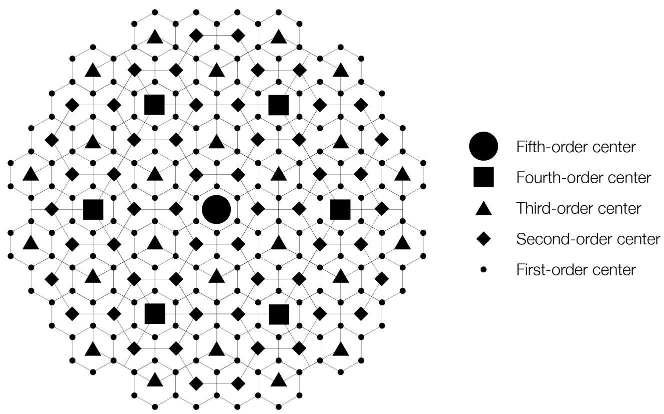


Fig. 1 Hexagonal spatial configuration of the five order central place system. Adapted from Openshaw and Veneris (2003).

Central Place Theory has been studied and evaluated from a number of perspectives since the 1930s. Berry and Garrison (1958b) performed a detailed analysis of central places in Snohomish County, Washington. They found the predictions of Central Place Theory largely confirmed by the data collected, with the exception of a small number of places that had seen a recent increase in population. They also used the data collected during this study to show that Lösch's idea of an economic equilibrium between places (Lösch, 1954) does not hold below certain population densities or levels of urbanization (Berry and Garrison, 1958a). Hsu (2012) later showed that the city sizes in Central Place Theory can be formalized using a power law model. Openshaw and Veneris (2003) evaluated the expected trip distributions in a central place model against spatial interaction models, finding that most spatial interaction models were unable to produce the trip distributions predicted by Central Place Theory.

2.2 Geographic analyses of the Web and Wikipedia

Both Wikipedia and the Web as a whole have been employed in a number of ways to answer different kinds of geographic research questions. Frankenplace (Adams and McKenzie, 2012) is an online application that extracts and analyzes qualitative geographic information from travel blogs, turning it into a data source for similarity-based place search. In previous work, we have demonstrated that the shapes of real-world features can be approximated based on geotagged photos on the Web (Keßler et al, 2009). Gao et al (in press) show how Volunteered Geographic Information can even be used for the construction of gazetteers. Salvini (2012) applies spatialization techniques to the English language edition of Wikipedia in order to analyze the functional structure of the global network of cities.

Likewise, the spatial aspects of Wikipedia have already been the focus of different studies. Takahashi et al (2011) present an approach to extract the significance of spatio-temporal events from Wikipedia. Their idea of links as impact propagation is similar to our approach that uses the number of references as an indicator of centrality of a place. Concerning the locations of contributors to Wikipedia, Lieberman and Lin (2009) show that the geographic coordinates of pages edited by a user often cluster tightly. Hecht and Gergle (2010) show, however, that this *localness* does not generally apply to any kind of Volunteered Geographic Information. Hecht and Moxley (2009) have conducted an extensive experiment to demonstrate the validity of Tobler's First Law of Geography (Tobler, 1970) in Wikipedia across different language editions of the online encyclopedia. They have shown empirically that places that are closer to each other in geographic space are also more likely to be related – i.e., interlinked – on Wikipedia.

In this paper, we take the concept of relatedness one step further by investigating how strong a place relates to another one, using the number of links and mentions as indicators. We show that this degree of relatedness reflects the functional dependencies between places as explained by Central Place Theory.

3 Data Access and Preprocessing

The dataset used in this study has been limited to a bounding box spanning the area between 50N, -128W (west of Vancouver Island, British Columbia, Canada) and 25N, -64W (north of Puerto Rico). It thus contains all of the contiguous United States as the focus area of our study. We limited the dataset under consideration to this area to have a consistent dataset in terms of language, taking into account only articles from the English language version of Wikipedia.[3] Previous research has shown that user generated content is not as local as the premise of Volunteered Geographic Information (Goodchild, 2007) suggests, especially in the case of Wikipedia (Hecht and Gergle, 2010). Nonetheless, using this combination of geographic area and Wikipedia articles in the main language spoken in this area should avoid the introduction of inconsistencies that arise from crossing language barriers. We hence defer the investigation of potential differences across geographic regions and language barriers to future research. Finally, this dataset is still tractable enough in terms of overall size in order to explore the feasibility of the general idea of this paper.

Since Wikipedia itself does not support queries by location through its API[4], we used DBpedia instead. DBpedia (Lehmann et al, 2012) provides facts extracted from Wikipedia as structured Linked Open Data (Berners-Lee, 2009). Among the facts extracted from Wikipedia are the geocordinates that are provided at the top right of a page for many subjects that have a geographic location. The coordinates are represented using the W3C Basic Geo Vocabulary (W3C Semantic Web Interest Group, 2004). This allowed us to retrieve all English language Wikipedia pages and their geographic coordinates within our bounding box, using DBpedia's SPARQL (Harris and Seaborne, 2013) endpoint.[5] The result was fed into a *places* collection stored in a local MongoDB[6] instance, consisting of entries of the following form:

---

[3] `http://en.wikipedia.org`

[4] `https://www.mediawiki.org/wiki/API`

[5] `http://dbpedia.org/sparql`

[6] `http://mongodb.org`

```
{ "_id"  : ObjectId("5466a15e080cb903020330fe"),
  "loc"  : { "type" : "Point",
             "coordinates" : [ -73.99028015136719,
                               40.62472152709961 ] },
  "page" : "http://en.wikipedia.org/wiki/Brooklyn" }
```


This collection of all georeferenced Wikipedia pages was used in the next step to download the actual contents of each page, using the XML export function of the Wikipedia API. The XML export is more straight-forward to parse than the actual HTML pages, while providing the same information. Each XML document was parsed for links to other georeferenced Wikipedia pages listed in our places collection, as well as for further *mentions* of such linked pages. As an example, when parsing the contents of the page for `http://en.wikipedia.org/wiki/Hoboken,_New_Jersey`, we will find links to the page for New York City. In Wiki syntax, this is represented as `[[New York City]]`, and rendered as

```
<a href="http://en.wikipedia.org/wiki/New_York_City">New York City</a>
```

by the MediaWiki engine driving Wikipedia. If we find such a link, we also scan the whole page for any other occurrences of the words *New York City* (without a hyperlink), since it is common practice to only link the first occurrence of a subject to its Wikipedia page, and not every single one. Taking these mentions into account gives us a more detailed impression of how often a georeferenced page is being referred to from other pages. Links of the form `[[Washington, D.C.|Washington]]`, that provide a different text to be shown (*Washington* in this example), are taken into consideration the same way.

By parsing all pages in this fashion, we built a collection of *links*, covering all pairs of pages that link to each other. One link consists of the linking page (*from*), the linked page (*to*), the number of actual *links*, the number of *mentions*, and the geographic *distance* (in meters) between the two georeferenced pages,[7] calculated from their respective coordinates in the *pages* collection:


```
{ "_id"      : ObjectId("54831bd6080cb9001a09ebb3"),
  "from"     : "http://en.wikipedia.org/wiki/Hoboken,_New_Jersey",
  "to"       : "http://en.wikipedia.org/wiki/New_York_City",
  "links"    : NumberLong(4),
  "mentions" : NumberLong(49),
  "distance" : 4618.070713194219 }
```


4 Dataset Characteristics

As of November 14, 2014, the dataset we retrieved in the way described in the previous section consists of 242,896 georeferenced subjects in the English language edition of Wikipedia. Parsing the pages' contents extracted 1,517,772 unique combinations of referring (*from*) and referred places (*to*), each with information on counts of links and mentions, and the distance between the respective pages as outlined in the previous section. The majority of those linking pages only contain a single link to the referenced page, as shown in Table 1. On average, a linking page contains ~ 1 link and ~ 0.86 textual mentions of the referred page. The maximum values observed are 28 links and 100 mentions, respectively, both of which are reached by multiple links. The average link is between places that are ~ 264km apart, whereas half of all links in our collection are between places that are ~ 21km or less apart. The biggest distance crossed by any of the links in our collection is ~ 5496km, which is about 600km short of the diameter of our bounding box.

[7] The distance has been calculated as great circle distance assuming a spherical Earth. The errors introduced by this simplification should be negligible in the context of this study.

Table 1 Overview of the link collection.

	Links	Mentions	Distance (meter)
Min	1	0	0
Median	1	1	21254
Mean	1.035	0.8554	264972
Max	28	100	5496277

Figure 2 shows the frequency distribution for the distances of each link. For this bar plot, the collection of all links has been divided into 100,000 subsets based on distance bins, each 54m wide.[8] Each bar in the chart shows the number of links in this bin, starting from the $0-54$m bin, at the very left, followed by the $55-109$m bin, etc. The orange bars show the number of pages linking to each other within the bin, while the gray bars are weighted by the number of references: For example, let place A and place B be 300m apart, and the page for place A links to the page for place B 3 times, and contains 2 more mentions. This would result in a single count in the $270-324$m bin for the unweighted bars (orange), and in 5 counts for the weighted bars (gray).

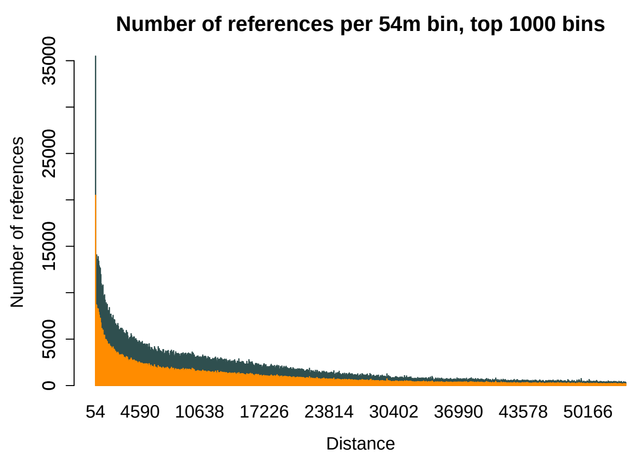


Fig. 2 Number of references in each 54m bin, weighted (gray) versus unweighted (orange). Note that these are *not* stacked to reflect that the unweighted links are included in the weighted references. The chart is limited to the top 1% of all bins, i.e., the rightmost bin ends at 54km.

The bar plot clearly shows the expected power-law distribution, i.e., places closer to each other link to each other more often, and they also *mention* each other more often in the text. Note that Figure 2 only shows the top 1% of the whole distribution, i.e., the number of links converges very quickly towards 1, and the number of mentions towards 0. We will make use of this fact in our detection of central places in Section 5. Figure 3 confirms this at the individual level, showing the total number of references (links and mentions) against the distance for all ~ 1.5 million links: Very high numbers of references can only be observed between places that are spatially close to each other, while the vast majority of links only contains a small number of references.

Table 2 shows the ten places with the highest number of incoming references, i.e, these are the most linked-to and mentioned places in our collection. Unsurprisingly, it consist of large-scale administrative units, led by the United States with close to to 200,000 incoming references. Several US states lag behind at about 30,000 references. The first cities in this

[8] While the choice of the number of bins is arbitrary, comparable results have been obtained with different bin sizes.

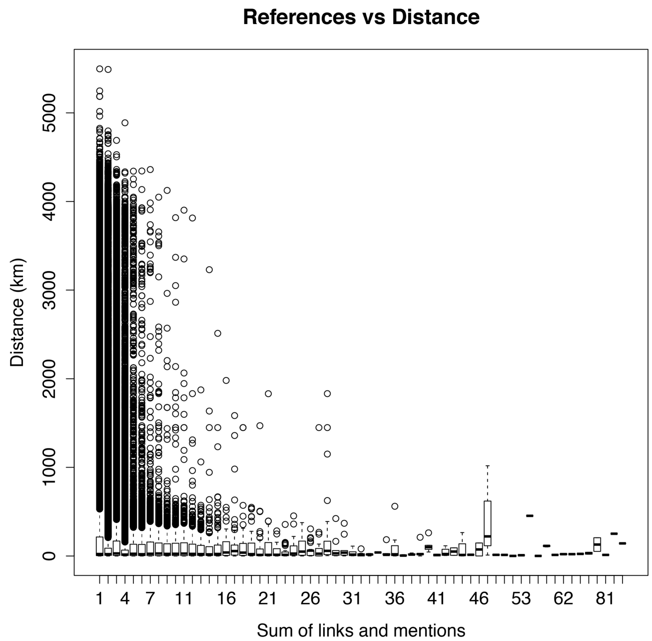


Fig. 3 Boxplot showing the variance in distances between places, grouped by total number of references (links and mentions) for all 1.5 million links.

ranking are New York City at rank 17 (8647 references), Chicago at rank 19 (6892 references), and Washington, D.C. at rank 24 (4718 references). While the order of this ranking is hardly surprising, it shows that these large-scale administrative units need special handling during our identification of central places.

Table 2 Overview of the top referenced places (links plus mentions).

	Place	**References**
1	http://en.wikipedia.org/wiki/United_States	199,605
2	http://en.wikipedia.org/wiki/California	35,990
3	http://en.wikipedia.org/wiki/Ohio	29,598
4	http://en.wikipedia.org/wiki/New_York	28,928
5	http://en.wikipedia.org/wiki/Illinois	27,600
6	http://en.wikipedia.org/wiki/Wisconsin	27,125
7	http://en.wikipedia.org/wiki/Indiana	26,893
8	http://en.wikipedia.org/wiki/Texas	25,132
9	http://en.wikipedia.org/wiki/Florida	20,981
10	http://en.wikipedia.org/wiki/Kentucky	17,776

5 Analyzing Central Place Structures

This section introduces an approach to analyze central place structures in Wikipedia based on the dataset discussed in Section 4. Using the four largest cities in the United States as case studies, we discuss the influence of weights and semantic aspects of the approach.

5.1 General approach

The premise of this research is that link structures in Wikipedia reflect real-world dependencies of smaller settlements (e.g., towns) on a central place (e.g., a city). We hence interpret every link as a pointer to a central place, where (a) the number of references in that specific link reflects the degree of dependency between the respective places, and (b) the total number of incoming references reflects the centrality – or relative importance – of a place. Following these assumptions, we can reveal the structure around a central place P as follows:

- Retrieve all links pointing to P.
- Remove all links that are beyond a weighted distance D, since many places have incoming links from places for away, where no spatial interaction in the sense of Central Place Theory is given (e.g., references between the pages of partner cities). This step is crucial and will be discussed in more detail in Section 5.2.
- Generate the convex hull of all remaining links to represent the range of P.
- For every remaining link, inspect the linking place and calculate its own relative importance based on its number of incoming links.
- Remove all links from this list where the linking place either has an administrative relationship to P, is not a settlement (see Section 5.3), or where the place has already been added to the structure of P in a previous iteration.
- From the remaining candidates, keep the top 6 closest places I according to their weighted distance. The number 6 follows from the hexagonal segmentation of space underlying Central Place Theory.
- Iteratively repeat the process for each place I to reveal the structure at the next lower order.

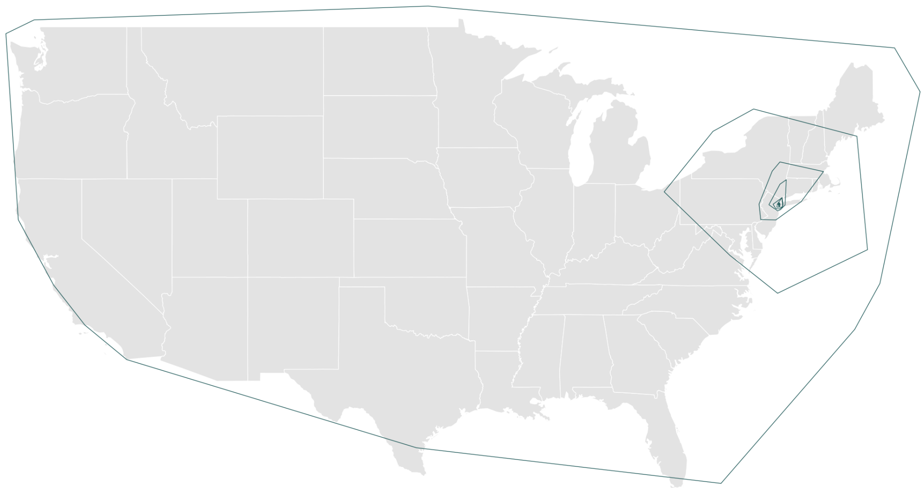
This approach will yield the 6 most relevant settlements at the next lower order; e.g., if P is a regional capital (fifth-order center), the first iteration will yield 6 cities (fourth-order center) in the range of P. The second iteration will yield the 6 most relevant towns (third-order centers) in range of each of those 6 cities, yielding a total maximum of 36 towns. Some of these 36 places will most likely appear twice at the same order, i.e., a third-order place may be in the range of two second-order places. Under perfect conditions, this would yield 24 unique third-order places, as every second-order place shares 4 third-order places in its range with another second order place (see Figure 1).

5.2 Distance and weighting considerations

Since many places in our collection receive incoming references from places all over the US, a mechanism is required to reliably generate a realistic convex hull representing the place's range. It needs to weigh the distance crossed by a link from a lower-order to a higher-order place against the degree of dependence. As discussed in Section 1, we use the number of references as an indicator of dependence, i.e., the higher the number of references, the more dependent a place is. We use a naïve approach here, where the weighted distance d_w for a link is defined as the geographic distance d divided by the number of references r:

$$d_w = \frac{d}{r}$$

This approach causes lower-order places with a high number of references to be drawn towards the higher-order place, figuratively speaking. If a place A is twice as far away as a place B from a higher-order place P, they would be assigned the same weighted distance if A has twice as many references to P as B.



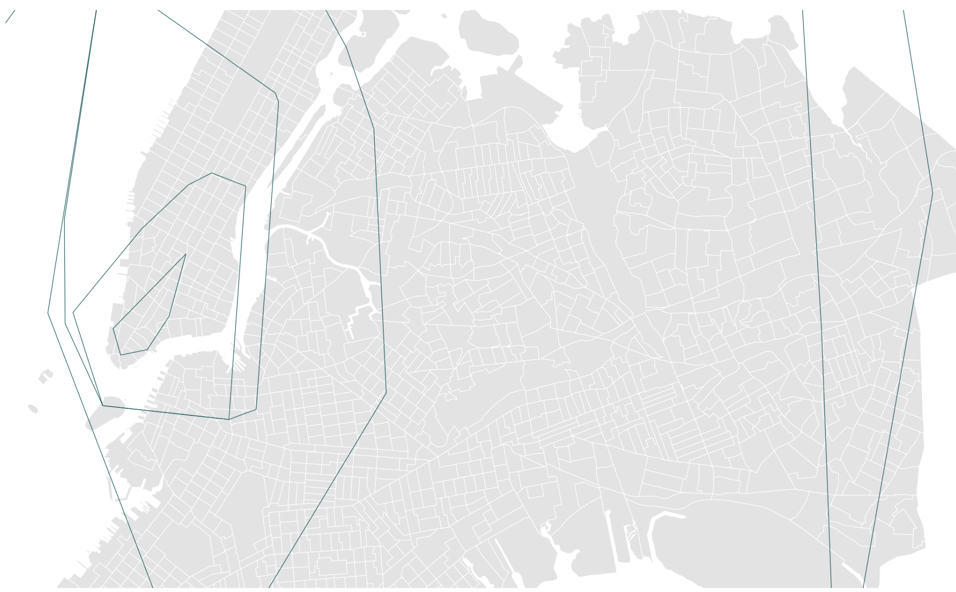


Fig. 4 Illustration of different quantiles of weighted differences. Every convex hull represents 10% quantiles of the incoming weighted links for New York City. When all links are taken into account, the convex hull contains the entire lower 48 states (top), whereas the 10% quantile contains only a few blocks in Lower Manhattan (bottom).

Using this weighting approach, we generate the input for the actual selection of the referring places we want to accept as being in the range of a place P. Figure 4 gives an overview of the 10% to 100% quantiles for the case of New York City, with 10% increments (the smallest area in the right part of Figure 4 is the convex hull for the 10% quantile, the next larger one for the 20% quantile, etc). We have experimented with different quantiles and found that taking into account the 75% quantile of all weighted links referring to a place yields the most realistic results in the case of the four metropolitan areas under consideration (see Section 6 for a more detailed discussion).

5.3 *Semantic aspects*

Large administrative areas tend to receive a high number of references, as shown in Table 2. Only relying on the 75% quantile of weighted distances would hence weave these administrative units into our central place structure. In most cases, these references would be meaningless for our purpose, though; the state of Illinois as a geographic entity does not contribute anything to the centrality of Chicago. It is rather the settlements *within* Illinois that interact with Chicago, and bear its importance in the central place structure. The same goes for counties and the federal state. Likewise, very small "places" – in the sense of some real-world entity that has a georeferenced Wikipedia page – are not meaningful in our structure. Parks, buildings, or companies should not be reflected in our structure, even if they are within a place's range, and their number of incoming references indicates centrality.

We make use of (a) the administrative place hierarchy and (b) information about the types of things we are looking at to decide whether to include a candidate in the central place structure or not. Places are only included in the structure if:

- They are at the same or a lower level in the administrative hierarchy as the place P under consideration, according to GeoNames.
- They are not a child of P in the administrative hierarchy, according to GeoNames (e.g., Brooklyn would be excluded if we are looking at New York City, as it is one of the city's five boroughs and hence a child in its hierarchy).
- They are of type settlement[9] (including any subtypes), as defined in the DBpedia ontology (Lehmann et al, 2012).

The following section evaluates the results obtained using our methodology for New York, Los Angeles, Chicago, and Houston.

6 Evaluation

This section evaluates the approach, looking at the results obtained for the four largest metropolitan areas in the US. For all figures in this section, black is used for fifth-order centers (i.e., New York City, Los Angeles, etc.), red for fourth-order centers, and orange for third-order centers. Meaningful results for the second- and first-order centers could not be obtained, due to the very low number of incoming links to the pages for the identified third-order centers.

Figure 5 shows the four central place structures evaluated, using the 75% quantile of the weighted links between places. The difference in scale between the four areas is evident. While we have not conducted a systematic evaluation against population density, it seems like structures yielded in higher population density areas are more compact. This finding needs to be confirmed by taking into account more examples in the future, and by evaluating the results against population density data.

Beverly, Massachusetts is incorrectly assigned as a third-order center to Beverly Hills, California, due to the high number of references to Beverly Hills on its page. This is an outlier that is not handled well by our approach yet. Beverly Hills, however, does show some other uncommon properties, namely that we could only identify 3 linking places outside of its own administrative hierarchy, one of them being Beverly on the East Coast. This low number of incoming links from the vicinity may point to a very separate, almost *gated* community. Independent of the underlying reasons, such outliers could be handled by a "hard" geographic cutoff distance, beyond which places are not taken into consideration

[9] See `http://dbpedia.org/ontology/Settlement`

any more. This would also speed up the computation of the central place structures when automating this approach further and expanding it to the whole world.

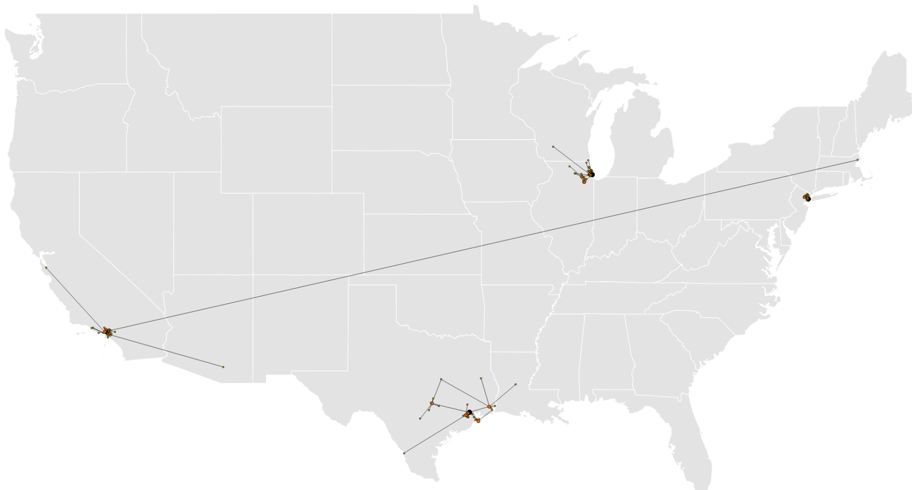


Fig. 5 Overview of the four central place structures evaluated, using the 75% quantile of the weighted links between places.

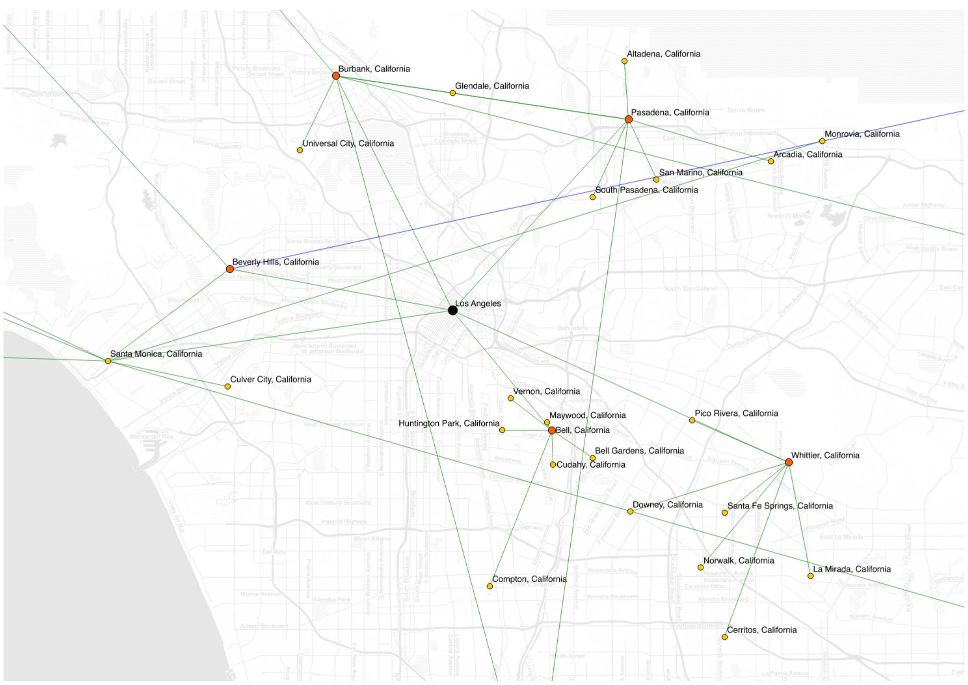


Fig. 6 Central place structure for Los Angeles.

All four structures face away from the water, which intuitively makes sense, but prevents a meaningful comparison with the structures predicted by Central Place Theory. Some of the fourth-level centers, however, approach a hexagonal configuration of space, such as

Bell and Pasadena in Figure 6, or Montclair in Figure 7. Using the underlying map as an indicator for population density again, the lower-order centers generally seem to lean towards areas with higher population areas. This is also to be expected, but needs further investigation. The administrative hierarchy of New York City also heavily influences the results shown. Since Queens, Brooklyn, and the Bronx were among the most central places linking to New York City, but excluded because of their being part of New York City, the whole central place structure faces towards New Jersey.

Fig. 7 Central place structure for New York City.

Both the structures for Houston and Chicago (Figure 8) show a small number of links that span long distances, which is somewhat unexpected given the high population densities in these areas. The number of references retrieved from Wikipedia indicate that many places in Texas are not as well documented as the New York City or Los Angeles areas, for example. This may explain these unexpected results, as our method strongly relies on a reasonably detailed input. As an example, when processing the structure around Bellaire, Texas, we only identified one third order place linking to it within the 75% quantile.

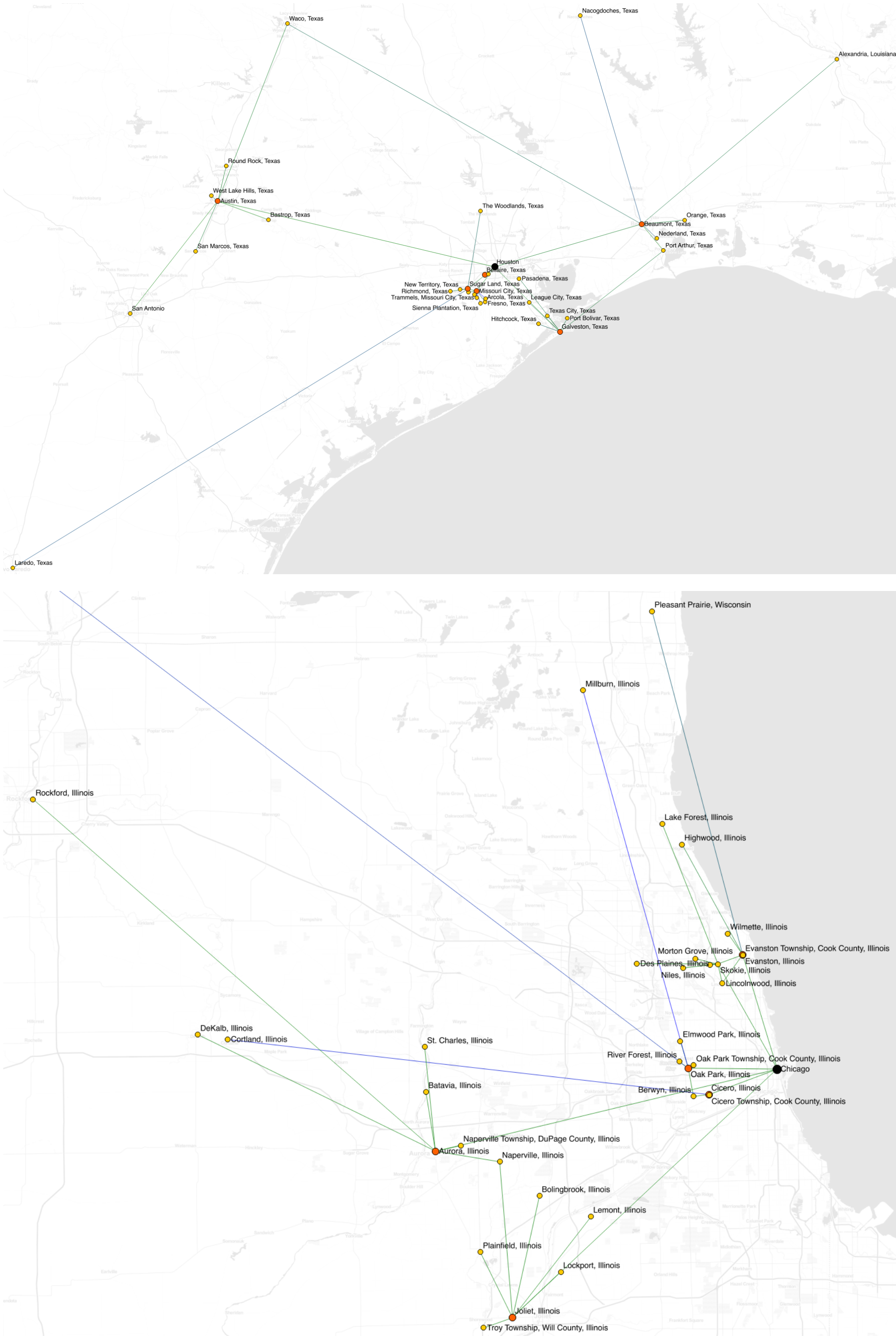


Fig. 8 Central place structure for Houston (top) and Chicago (bottom).

7 Conclusions

We have introduced an approach to extract the structure around a central place and its range from the link structure in the English language edition of Wikipedia. Using weighted distances and semantic annotations, we have demonstrated that the Web and its hyperlink structure can indeed be used to infer spatial structures in the real world. While the results vary significantly depending on population density and natural features – all places considered in the paper are near the sea or a large lake –, parts of the identified structures match the predictions of Central Place Theory well. The presented results are only a first indication that the Web does not only exhibit patterns of spatial autocorrelation (Hecht and Moxley, 2009) and the shapes of real-world features (Keßler et al, 2009), but it also reflects interactions between places. The study indicates that the link structure on the Web mirrors which places functionally depend on each other, and to what degree. Our results also point to the fact that the link structure in Wikipedia is only useful down to the third-order centers, as these are usually already small towns whose Wikipedia pages do not have any significant numbers of incoming links.

While the structures around the regional capitals investigated in this paper seem intuitive, the results clearly need a more thorough, quantitative analysis, also in order to fine-tune the process of identifying the structures. With the software tools built for this process, the next step will be to fully automate the generation of central place structures from Wikipedia. This will allow us to experiment with different variations of the approach, and run it on a larger input body.

Acknowledgements The base maps used in Figures 6–9 have been provided by Stamen Design[10] under a Creative Commons License.[11] The maps are based on OpenStreetMap data, provided under the Open Database License.[12]

References

Adams B, McKenzie G (2012) Frankenplace: An Application for Similarity-Based Place Search. In: Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media

Baskin CW (1966) Central Places in Southern Germany. Prentice Hall

Berners-Lee T (2009) Linked Data – Design Issues. Online: `http://www.w3.org/DesignIssues/LinkedData.html`

Berry BJL, Garrison WL (1958a) A Note on Central Place Theory and the Range of a Good. Economic Geography 34(4):304–311

Berry BJL, Garrison WL (1958b) The Functional Bases of the Central Place Hierarchy. Economic Geography 34(2):145–154

Brush JE (1953) The Hierarchy of Central Places in Southwestern Wisconsin. Geographical Review 43(3):380–402

Christaller W (1933) Die zentralen Orte in Süddeutschland. Gustav Fischer, Jena

Gao S, Li L, Li W, Janowicz K, Zhang Y (in press) Constructing gazetteers from volunteered big geo-data based on Hadoop. Computers, Environment and Urban Systems

Goodchild M (2007) Citizens as sensors: the world of volunteered geography. GeoJournal 69(4):211–221

Harris S, Seaborne A (2013) SPARQL 1.1 Query Language. W3C recommendation available from `http://www.w3.org/TR/sparql11-query/`

[10] `http://stamen.com`

[11] `http://creativecommons.org/licenses/by/3.0`

[12] `http://www.openstreetmap.org/copyright`

Hecht B, Moxley E (2009) Terabytes of Tobler: Evaluating the First Law in a Massive, Domain-Neutral Representation of World Knowledge. In: Hornsby K, Claramunt C, Denis M, Ligozat G (eds) Spatial Information Theory, Lecture Notes in Computer Science, vol 5756, Springer Berlin Heidelberg, pp 88–105, DOI 10.1007/978-3-642-03832-7_6, URL http://dx.doi.org/10.1007/978-3-642-03832-7_6

Hecht BJ, Gergle D (2010) On the localness of user-generated content. In: Proceedings of the 2010 ACM conference on Computer supported cooperative work, pp 229–232

Hsu WT (2012) Central Place Theory and City Size Distribution. The Economic Journal 122(563):903–932

Keßler C, Maué P, Heuer JT, Bartoschek T (2009) Bottom-Up Gazetteers: Learning from the Implicit Semantics of Geotags. In: Janowicz K, Raubal M, Levashkin S (eds) Third International Conference on GeoSpatial Semantics (GeoS 2009), Springer Lecture Notes in Computer Science 5892, pp 83–102

Lehmann J, Isele R, Jakob M, Jentzsch A, Kontokostas D, Mendes PN, Hellmann S, Morsey M, van Kleef P, Auer S, Bizer C (2012) DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. Semantic Web Journal 1:1–5

Lieberman MD, Lin J (2009) You Are Where You Edit: Locating Wikipedia Contributors through Edit Histories. In: Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009, San Jose, California, USA, May 17–20, 2009

Lösch A (1954) The Economics of Location. Yale University Press, New Haven, CT

Openshaw S, Veneris Y (2003) Numerical experiments with central place theory and spatial interaction modelling. Environment and Planning A 35(8):1389–1404

Takahashi Y, Ohshima H, Yamamoto M, Iwasaki H, Oyama S, Tanaka K (2011) Evaluating significance of historical entities based on tempo-spatial impacts analysis using wikipedia link structure. In: Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia, pp 83–92

Tobler WR (1970) A computer movie simulating urban growth in the detroit region. Economic Geography 46:234–240

W3C Semantic Web Interest Group (2004) Basic Geo (WGS84 lat/long) Vocabulary. Online: http://www.w3.org/2003/01/geo/

Salvini M (2012) Spatialization von nutzergenerierten Inhalten für die explorative Analyse des globalen Städtenetzes. PhD thesis, University of Zurich