The role of cultural importance in sister city relationships

Extended Abstract

Mridul Seth*
BITS Pilani KK Birla Goa Campus
Goa, India
seth.mridul@gmail.com

Jeŕom̂e Kunegis
University of Namur
Namur, Belgium
jerome.kunegis@unamur.be

Renaud Lambiotte
Mathematical Institute, University of Oxford
Oxford, UK
renaud.lambiotte@maths.ox.ac.uk

ABSTRACT

The modern concept of sister city is where two communities, cities or provinces engage in mutual social and cultural exchanges. With more than 20,000 such relationships around the world we examine a model of cultural importance between these cities. Cultural importance is a subjective term and to look at it from a quantitative point of view we use the data available on Wikipedia. We look at the number of languages of Wikipedia articles, page views, population and content length. We examine the role of cultural importance as a driving force in the sister city network and measure the assortativity coefficient of the network as a function of cultural importance.

KEYWORDS

Urban Networks, Wikipedia, Assortativity, Sister City

ACM Reference format:
Mridul Seth, Jeŕom̂e Kunegis, and Renaud Lambiotte. 2018. The role of cultural importance in sister city relationships. In *Proceedings of WikiConference, Paris, France, April 2018 (WWW'18),* 2 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Sister city relationships have been studied from a historical and geographical perspective by Zelinsky in his paper *The Twinning of the World* [4]. He argues that there is an unwritten rule that the two places should be roughly comparable in size and must be compatible partners and we look at the compatibility between cities from a cultural importance perspective in this paper. To study this network of relationships we create a network out of this data where the cities are represented by nodes and the edges represent their sister city relationships. A network representation of sister cities was previously done by Kaltenbrunner et al in their work *Not all paths lead to Rome: Analyzing the network of sister cities* [1] where they created a network from Wikipedia entries of sister cities and analyzed the network structure and found out that there is negligible influence of geographical distance between sister cities.

In this paper we look at the role of cultural importance of sister cities in the sister city network, inspired from the Historical Popularity Index [3], and discuss the driving force behind these sister city relationships.

We extract the network using the SPARQL query service of Wikidata and find 14,329 nodes and 20,738 edges (as of 15th January 2018). The nodes are not necessarily just cities, these could be small villages, towns even counties and states represented by the term "Human Settlements" internally in Wikidata. We create an undirected and unweighted network out of this dataset. To collect metadata about the cities we again use Wikidata and find the number of languages, population, page length in every language, page views in every language.

2 CULTURAL IMPORTANCE

Cultural importance of a place or city is a subjective term and to look at it from a quantitative point of view comes with its own challenges, but today with the wealth of data available, we can try to create a model which gives us a clearer picture.

A basic way to measure the importance of any article(not just cities) on Wikipedia is to look at the number of languages the article is available in, for example the article on Washington, D.C is available in 218 languages, Zurich in 137, Eschwege (town in Germany) is 33. This observation is consistent with our expectations that big cities and capitals are more important than other cities in the global network. To collect this data we scrap the Wikidata entries of the cities that are present in our sister city network and find the number of wikipedia.org sites, i.e the number of languages the article about the city is available in.

$$L_i = \#\text{ of Wikipedia articles available} \tag{1}$$

Popularity of an article also gives us an idea of importance of that article and to measure the popularity we look at the number of page views on the Wikipedia page. To avoid biases like spikes due to current events we collect the data from 1st July 2015[API restricted] to 1st January 2018 and find the average monthly page views. We do this for all the languages available for an article. While collecting this data we already have added a bias, as by just comparing the magnitude of number of page views of English Wikipedia of an Italian city and an American city doesn't gives an accurate image because its possible that an Italian city has more page views in Italian than English. Hence we get a skewed view of popularity. To take this into account we measure the absolute maximum for a city irrespective of the language.

*This work was done while visiting University of Oxford and Univeristy of Namur

WWW'18, April 2018, Paris, France

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

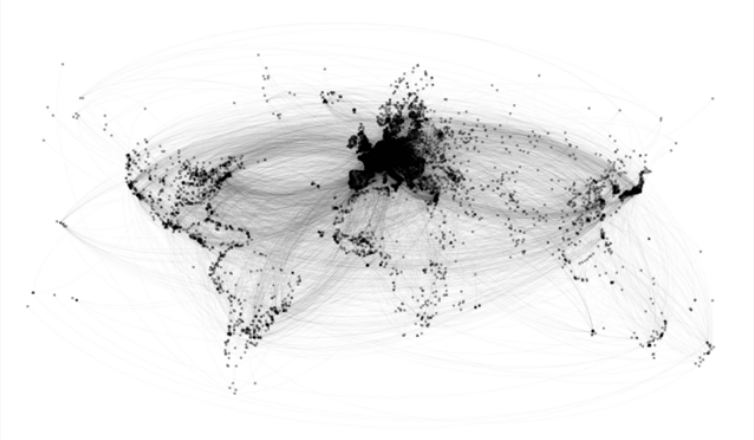

Figure 1: Worldwide sister city network

$$PV_i = \max(\text{page views of all languages for city i}) \quad (2)$$

Another way to look at the popularity is to look at the content length of articles in different languages. There are a lot of articles which are in depth detail in one language, for example the content length(as returned in HTTP headers) for the article on Lausanne is '68178' in English, it is '16126' in Marathi (a regional Indian language). We use this information in addition to the information of page views to get a better holistic overview of importance in our sister city network, and just like page views we take the absolute maximum to avoid creeping in a bias.

$$C_i = \max(\text{content length of all languages for city i}) \quad (3)$$

We define cultural importance of a city, i by

$$CI_i = H(L_i) + H(PV_i) + H(C_i) + Sgm(500 * H(POP_i)) \quad (4)$$

where POP_i is the population of the city i.

$$Sgm(K * X) = \frac{1}{1 + \exp(-K * X)} \quad (5)$$

$$H(X_i) = \ln(1 + \frac{X_i}{max(X_{ic})}) \quad (6)$$

where $max(X_{ic})$ is the maximum magnitude of X characteristic in the set of all cities.

3 THE DRIVING FORCE

The assortativity coefficient [2] of a network gives us information about the connections between similar nodes (or dissimilar nodes). A high and positive coefficient, i.e. closer to 1 indicates that edges are between similar nodes and a negative coefficient, i.e -1 indicated that edges are between dissimilar nodes. A coefficient of 0 means that the network is non-assortative.

In the case of sister city network, the assortativity of a network can tell us about the driving force of sister city relationships. In the work [1], they found out degree assortativity to be 0.3407. Which means that just degree of a city is not a good enough metric for cities to seek relationships. In our network we find degree assortativity to be 0.385.

This is where the cultural importance of a city comes into the picture. The assortativity coefficient with respect to cultural importance for this network is 0.689. To make sure we haven't over fitted our model to the data set, we test this on different sub graphs of network and find consistently high assortativity coefficient between 0.65 - 0.75.

4 DISCUSSION

In this work we find out that cities with similar cultural importance tend to form sister city relationships. To further examine our model we create sample test cities which have characteristics of a cultural important city like high number of languages, page views and connect it with other culturally important cities to find an increase in the assortativity coefficient as expected and a decrease in assortativity coefficient for connecting high cultural importance cities with low cultural importance cities. This observation follows Zelinsky's argument that cities should be compatible partners, and in this case we use cultural importance as a measure of compatibility.

This is an extended abstract and work in progress. This work is intended to be a full paper.

REFERENCES

[1] Andreas Kaltenbrunner, Pablo Aragón, David Laniado, and Yana Volkovich. 2013. Not all paths lead to Rome: Analysing the network of sister cities. In *International Workshop on Self-Organizing Systems*. Springer, 151–156.

[2] Mark EJ Newman. 2003. Mixing patterns in networks. *Physical Review E* 67, 2 (2003), 026126.

[3] Amy Zhao Yu, Shahar Ronen, Kevin Hu, Tiffany Lu, and César A Hidalgo. 2016. Pantheon 1.0, a manually verified dataset of globally famous biographies. *Scientific data* 3 (2016).

[4] Wilbur Zelinsky. 1991. The twinning of the world: sister cities in geographic and historical perspective. *Annals of the Association of American Geographers* 81, 1 (1991), 1–31.