Decision Support Systems xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Decision Support Systems

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

Geography of online network ties: A predictive modelling approach

Swanand J. Deodhar *, Mani Subramani, Akbar Zaheer

Carlson School of Management, University of Minnesota, 321 19th Avenue South, Minneapolis, MN 55455, United States

ARTICLE INFO

Article history:
Received 20 September 2016
Received in revised form 24 April 2017
Accepted 4 May 2017
Available online xxxx

Keywords:
Geography
Online network ties
Psychic distance
Predictive modelling
Cluster analysis

ABSTRACT

Internet platforms are increasingly enabling individuals to access and interact with a wider, globally dispersed group of peers. The promise of these platforms is that the geographic distance is no longer a barrier to forming network ties. However, whether these platforms truly alleviate the influence of geographic distance remains unexplored. In this study, we examine the role of geographic distance with machine learning approach using a unique dataset of the network ties between traders in an online social trading platform. Specifically, we determine the extent to which, compared to other types of distances, geographic distance predicts the occurrences of the network ties in country dyads. Using cluster analysis and predictive modelling, we show that not only the geographic distance and network ties exhibit an inverse association but also that geographic distance is the strongest predictor of such ties.

1. Motivation for the research

Since the inception of internet, the world is moving towards a scenario in which individuals, irrespective of their location, can interact with a large, globally dispersed networks of peers, leading to transformative economic activities such as user innovation [26], online labor markets [9], and crowdsourcing [22]. Internet platforms drive this transformation by providing means to access and communicate with one's peers at practically non-existent cost. For such online intermediaries, the ideal world is the one in which individuals can engage in frictionless interactions and create network ties, rendering geographic distance inconsequential [11]. Although the idea of frictionless interactions is appealing, there is surprisingly little prior empirical research that informs the fundamental question – what is the role geographic distance in shaping online behavior?

A few recent studies have examined this issue [10] suggesting that distance adversely influences frequency and magnitude of dyadic ties in online context as well. Our study extends this literature in at least two ways. First, we isolate the extent to which geographic distance predicts the occurrences of dyadic ties by comparing its predictive power with that of the competing distance measures. Such an approach is necessary because a standard online platform does not directly provide geographic distance as an information cue to its users. Instead, it makes each user's nationality and other location information visible to other users. Therefore, geographic distance is only one of the several distance measures that can predict user's behavioral response. Any assessment of geographic distance as a predictor of network ties in an online context is incomplete without the inclusion of other forms of distances. Second, we examine geographic distance in a context that does not follow the two-sided platform structure which is predominant in the extant literature on distance effects in online settings. This difference is relevant to the occurrence of dyadic ties because on two-sided platforms dyadic ties are typically "cross-sided." Instead, our setting allows any user to form a tie with any other user on the platform, broadening the possible pool of users with whom ties can be established. In sum, our primary research question is as follows: *in a globally distributed network of individuals, in which users can create ties with any other user, whether and to what extent geographic distance predicts the occurrence of dyadic ties*?

We address this research question by using a dataset of dyadic ties obtained from an electronic investment platform, which we refer to as XTrader. The platform is meant for the currency and commodities trading and has a user-base of over a million traders, representing nearly 100 countries. XTrader is an appropriate choice for our study for several reasons. First, the platform allows traders to form direct ties with each other. Because all the traders are engaged in the same activity (i.e. trading), there are no distinct sides to the platform. Hence, every trader can form a tie with every other trader. Second, a trader can create a tie only by allocating a certain portion of their fund to the other trader. That is, each tie that a trader creates has a cost associated with it, allowing us to consider the existence of a tie as a conscious decision on a trader's part for which the trader is likely to consider available information cues about a potential tie partner. Third, the platform provides each trader's country as the only demographic information cue. This cue is publicly visible to everyone. The salience of trader's nationality enables the distance mechanism to come into play.

* Corresponding author.
E-mail addresses: deodh009@umn.edu (S.J. Deodhar), subra010@umn.edu (M. Subramani), azaheer@umn.edu (A. Zaheer).

http://dx.doi.org/10.1016/j.dss.2017.05.010
0167-9236/

Table 1
Summary of articles on distance effect in online contexts.

Study	Distance measure	Context	Key finding related to distance effect
Blum & Goldfarb [6]	1. Geographic distance, 2. Difference in GDP	Website visits	In taste-based digital products (e.g. music, games) users from a given country make significantly more visits to websites from geographically nearer countries.
Gefen & Carmel [18]	1. Geographic distance	Online labor markets (Rent A coder)	Except for the American clients, others tend to award project a geographically proximal agent. However, the preference gets mitigated if the agent and the client come from English-speaking countries.
Hortaçsu, Martínez-Jerez, & Douglas [21]	1. Geographic distance	Online auction sites (MercadoLibre and eBay)	Buyers prefer to buy from sellers who are geographically proximal, preferably, within a driving distance or same city.
Agarwal, Catalini, & Goldfarb [3]	1. Geographic distance	Crowdfunding (Sellaband)	Geographically proximal investors are likely to fund a project early while the distant ones invest once the project accumulates investments
Takhteyev, Gruzd, & Wellman [41]	1. Geographic distance 2. Language similarity 3. Air travel (number of direct flights between two locations)	Microblogging (Twitter)	Geographic distance has an adverse influence on the number of Twitter ties. However, it is not robust to the magnitude of air travel between two locations.
Burtch, Ghose, & Wattal [10]	1. Geographic distance 2. Cultural distance	Crowdfunding (Kiva.com)	Both geographic and cultural distance significantly reduce the number of ties between donors and receivers The two distances are substitutes of each other
Lin & Viswanathan [27]	1. Geographic distance	Crowdfunding (Prosper.com)	Donors are more likely to donate money to those who are geographically proximal even when such a donation is not in the donor's economic interest
Posegga, Zylka, & Fischbach [32]	1. Geographic distance	Crowdfunding (Kickstarter)	While there exists gender-based (male(female) investors investing money in projects started by male(female) entrepreneurs), geographic distance does not matter.
Lengyel et al. [25]	1. Geographic distance between towns	Intra-Country Social Network in Hungary (International Who is Who)	Individuals are more likely to form links with those who are from geographically proximal towns. However, in the online social network, the probability of a tie's existence decays at a slower rate with respect to distance.

We adopt the data construction similar to Burtch et al. [10]. The dataset we obtained from XTrader pertains to trader-dyads observed for 46 consecutive weeks. Because we are interested in predicting the number of ties between pairs of countries, we begin by aggregating these observations to country-dyads. In the final dataset, we have 266,570 data points consisting of 5795 unique pairs observed over 46 weeks. These pairs represent 95 distinct countries. In each country-dyad, the outcome variable is the count of distinct ties from one country to another, and the predictors are different types of distances between the countries in that particular dyad. More specifically, we use psychic [13] and geographic distance measures.

We employ cluster analysis and predictive modelling because we are interested in assessing the extent to which distances and specifically, geographic distance predicts the occurrences of ties. First, using cluster analysis, we obtain a robust, 4-cluster solution. Among the 4 clusters, one cluster represents country dyads that have the highest geographic

Table 2
Variable descriptions and summary statistics.

Variable name	Brief description	Source	Mean	Standard deviation
Count of ties in Country Dyads	The number of unique ties from the source country to the destination country	Obtained from the platform	237.24	176.37
Geographic distance	The distance (in KM) between two most populated agglomerations. The distance is computed using the latitudes and longitudes of the agglomerations	Mayer & Zignago [28]	7216.27	4831.15
Social (political)[a]	The difference in the political ideology scale for the two countries	Dow & Karunaratna [14]	0.41	0.27
Democracy (political)[a]	The composite score that captures the difference between two countries on a series of scales including POLICON, Political rights scale, civil liberty scale, and POLITY V scale.	Dow [13]	0.56	0.48
Religion	A composite score capturing the extent to which: • The major religions of the two countries are different • One country practices the religion of the other country		0.79	0.44
Language	A composite score capturing the extent to which: • The national languages of the two countries are different • The population in one country speaks the national language of the other country		1.51	1.36
Education	The composite score for the • The difference in the literate adults • Population enrolled in the second as well as the third levels of education		0.79	0.58
Industrial Development	The difference in industrial development between two countries. Dow [13] measures industrial development using ten sub-factors • GDP difference • Ownership of electronic goods (TV, radio, telephones) and car • Energy consumption • Newspaper usage • Percentage of the URBAN population		0.82	0.58

[a] For the original references of each scale, see Dow [13].