

A Cross-Country Investigation of User Connection Patterns in Online Social Networks

Christine Bauer
Johannes Kepler University Linz
christine.bauer@jku.at

Markus Schedl
Johannes Kepler University Linz
markus.schedl@jku.at

Abstract

Given the global expansion, the borderless nature, and the social impact of social media, this paper provides an examination of users' connection patterns in online social networks, more specifically the users' cross-country connection patterns. We study three highly different social media platforms, Facebook, Last.fm, and 500px, and approach two main research questions:

First, we set out to answer which countries' social media users are mainly connected with users within their own country; and which countries are characterized by a wide spectrum of cross-country (transnational) user connections. In doing so, we also identify the "attractor" countries, being characterized by alluring a large portion of users from other countries to connect to users in the respective attractor country. Second, we compare the results between the three social media platforms under investigation and analyze and discuss differences in the cross-country connection patterns. Third, we investigate whether countries' attractor values are correlated with cultural features (according to Hofstede).

Our results contribute to understanding the complex social ties between people and how they are reflected in connection behavior on social media.

1. Introduction

Long before it actually happened with the establishment and popularization of the Internet, Marshall McLuhan [1, 2] predicted in the 1960s that the world would be embraced by an electronic extension of our central nervous system that would be part of our popular culture. In his work, McLuhan envisioned and coined the concept of the *global village*: "The new electronic interdependence recreates the world in the image of a global village."

Today, online social networks (OSN)—such as Facebook, Instagram, or Twitter—have become

important means for global social exchange. Every day, millions of people worldwide send, receive, post, view, and like billions of messages via OSN. As with the borderless nature of the Internet, user connections in OSN are not limited to any single country and appear universal in scale [3]. So, has the *global village* become a reality in OSN?

Social connectedness in OSN across the globe has repeatedly been a target of research. In the early days of OSN, some authors (e.g., [4]) conjectured that the use of OSN may follow the spatial, structural, and cultural perimeters of the societal system that users belong to in the real world. Later studies (e.g., [5]) revealed that offline social connectedness and connectedness on OSN were separate constructs. In other words, OSN could allow to go beyond such spatial boundaries in the real world.

As evidenced, for instance, in a study comparing Korean and American users, people's motivation to engage in an OSN and to connect with other people varies across countries [6]. Therefore, a widespread approach to measure transnational social connectedness in OSN is by studying cross-country user connections and compare several countries (e.g., [7, 8, 9]). One limitation of existing research is that a large majority of studies on connection patterns focuses on Facebook (e.g., [7, 9, 8]). However, some authors (e.g., [10]) emphasize that the nature of user connections may also vary from OSN to OSN. Another limitation is that research on social connectedness in OSN tends to compare a small set of countries only (e.g., two countries in [11], three in [12]). Work that compares a large set of countries either focuses on the structure of the social graph (e.g., [9]) or discusses connections only on the level of the individual for a small set of users (e.g., [7]).

Against this background—and in contrast to above described works—, we delve into detail with respect to the bidirectional user-to-user connections aggregated on a country level and we consider an aggregation of individual cross-country connections.

In this investigation, we specifically approach the

following research questions:

- RQ1: What user connection patterns exist across countries? More specifically:
 - RQ1a: Which are the countries whose users have mainly intracountry user connections (i.e., connect with users within their own country)? Which are the ones that show transnational connection behavior?
 - RQ1b: What are the most important “attractor” countries, i.e., countries whose users are substantially more often the target of a friendship connection than other countries (in relative numbers)?
 - RQ1c: Is country attractiveness correlated with cultural aspects?
- RQ2: Are connection patterns (intracountry vs. transnational) comparable between different OSN?

Addressing these research questions, we present an analysis of cross-country user connections of three different OSN in this work: Last.fm¹—an OSN for music enthusiasts, 500px²—an OSN for photography, and Facebook³—a general-purpose OSN.

The results of this work contribute (i) to answering the general question whether the global village has become a reality in OSN and (ii) to a better understanding of user connection patterns in OSN—and, thus, deliver contributions to the core research area of social network analysis. Furthermore, this work has practical implications for the research field of personalized systems and recommender systems because it allows for advancements in *user modeling*. In particular, our results could help alleviating the new-user cold start problem in recommender systems. A new-user cold start problem occurs for users new to the system under consideration, because there is no user history about him or her and the system does not know the personal preferences (i.e., has yet created a user model). Creating a user model based on the country-specific approximations of tie strengths of user connections and the country’s attractor status would address this problem and could substantially alleviate the new-user cold start problem. More concretely, today’s systems frequently use single sign-on buttons, which allow new users to register with their Facebook, Twitter, or other OSN accounts, giving the system access to their profile information. Since user profiles commonly contain country information, in the absence

¹<https://www.last.fm>

²<https://500px.com>

³<https://www.facebook.com>

of item interaction data, our results could help trigger initial recommendations based on the typical connection patterns of users in the target user’s country.

The remainder of this work is organized as follows: Section 2 puts our work into context and presents related literature. Section 3 elaborates on the methodology we adopt to address the research questions. Furthermore, we describe the rationale for OSN selection and the respective approaches for data acquisition. In Section 4, we present the results of our study. As our analysis shows, despite global interconnections on a wide basis, for some countries we identify that their users stick with their kind. In addition, choosing three different OSN allows not only for cross-country analyses but also comparisons between general-purpose (Facebook) vs. specialized OSN (Last.fm/music, 500px/photography). As our results will show, there are specific differences. In Section 5, we discuss the practical implications of our results for personalized systems. Finally, we round off our work by a conclusion and outlook to future research in Section 6.

2. Related Work

2.1. Cross-country Comparisons of User Connections

A typical characteristic of social media is that users may connect with each other. Thereby, user connections in OSN may be unidirectional or bidirectional. For instance, “friendship” connections on Facebook, where both users have to agree to the connection, are bidirectional, leading to a symmetric social graph. The “follower” connections on Twitter, where mutual following is optional, are unidirectional, leading to an asymmetric graph of connections [12, 13]. While most studies on user connections focus on bidirectional connections (e.g., [9, 8, 14]), some studies (e.g., [12, 13]) analyze asymmetric “follower” connections.

As a research focus, many studies compare the social graph of connections of users in one country with those of another country. For instance, user connections in the United States versus in Korea are examined for their strength and concentration in [11]. The authors found that users in the United States generally maintain larger but looser networks with a far greater portion of weak ties, whereas their Korean counterparts maintain smaller and denser networks with a roughly even ratio of strong and weak ties. Average path lengths within the country-specific social graphs were also found to vary across countries (for Brazil, Japan, and the United States) [12]. Typically, studies analyzing country-specific differences rely on one specific OSN,

for which they analyze various characteristics across two or more countries.

2.2. Approaches to Investigate Cross-country User Connections

Various approaches have been pursued to study transnational user connections across several countries. For instance, connection patterns of individual OSN users may be examined [7] or an aggregation of individual cross-country connections per country may be used to measure the relational strength between countries [9, 8]. Similar approaches can also be taken to analyze connections across regions within one country. For instance, user connections across counties within the United States—but also connections between those counties and foreign countries—have been studied based on an aggregation of connections [8].

Our work relies on and differs from related work in the following ways: We adopt an approach similar to [8] (i.e., aggregation of user connections). While [8] analyzed cross-county connections within the United States (and only some additional connections to foreign countries), we in contrast analyze within-country and cross-country user connections on a global scope. We consider a large set of countries, whereas related work usually compares two (e.g., [11]) or three (e.g., [12]) countries. Work comparing a large set of countries, either focuses on the structure of the social graph such as in [9] where the authors identified geographic and linguistic clusters bridging the country-specific social graphs, or uses an approach based on individual users' connection patterns (e.g., [7]). In contrast to these works, we delve into detail with respect to the bidirectional user-to-user connections aggregated on a country level and we consider an aggregation of individual cross-country connections. In addition, we compare three different OSN, one general purpose (Facebook) and two used by certain communities (Last.fm and 500px).

2.3. Dimensions of Cross-cultural Differences

When comparing user behavior in various countries, also cultural aspects come into play. The concept of *culture* is broadly defined as the collective norms, rules, and values which distinguishes one group of people from another [15]. There are several approaches to define culture on a country level (i.e., national culture). The arguably most frequently cited and widely adopted framework to differentiate between cultures on a country level was presented by Hofstede [15, 16]. He proposed six⁴ distinctive dimensions to describe a

⁴Originally four; two were added later.

country's culture, i.e. individualism/collectivism, power distance, uncertainty avoidance, masculinity/femininity, and long-term orientation, and indulgence. We will use this framework to address RQ1c, thus, for investigating the relationship of a country's "attractiveness" for user connections and the cultural dimensions.

3. Methods and Materials

In this section, we first describe the rationale for the selection of OSN for our study (Section 3.1). Then, we elaborate on the approach taken to assess the cross-country connections between users in the selected OSN (Section 3.2) and describe the datasets used for our analysis (Section 3.3). This section rounds off with a pointer to the data we used for our analyses with respect to the cultural dimensions by Hofstede (Section 3.4).

3.1. Selection of Online Social Networks

In order to assess the cross-country relationships between users, we investigate the user connection information for Last.fm, 500px, and Facebook.

We chose two different topic- or community-specific OSN, i.e., Last.fm and 500px, for this analysis for the following reasons: Last.fm is very popular among music listeners. It accumulates the music taste of its users over the lifetime of their accounts. 500px, in contrast, is a social media platform for photographers, which is popular among amateur photographers and professionals alike. Both social media platforms only allow to create user connections that are bidirectional or commutative, i.e., if a user u_1 is connected to u_2 , user u_2 is also connected to u_1 , which makes the structure of connections between the two platforms comparable.

In addition, we analyze a general purpose OSN to be able to compare the results of the topic-specific OSN to those of a general one. This also provides a good basis for assessing the stability of the results. We chose Facebook because, among OSN, Facebook is the general purpose network that reaches the highest number of active users on a daily basis [17]. Also, Facebook mainly serves as a platform for real-world friends and acquaintances to interact online, and people usually only add connections on Facebook to individuals whom they know in the real world [18, 19, 20].

3.2. Approach

In order to assess the cross-country relationships between users in OSN, we investigate the user connection information for the three selected social media platforms, i.e. Last.fm, 500px, and Facebook. We take a similar approach to calculate these relationships

for pairs of countries as [8] do for pairs of countries within the United States.

More precisely, for each country c , the share of user connections maintained with other users in c is compared to the share maintained with users from other countries in the following way: For each pair of countries, c_1 and c_2 , we compute the share of users in c_1 that are connected to users in c_2 . This yields a (per-row) *normalized country connection matrix* (cf. Table 2, Table 3, and Table 4), based on which we will answer RQ1a.

For our analysis, we consider the top 20 countries in terms of total number of users based on the Last.fm dataset (cf. Section 3.3.1). We take Last.fm as the basis because—unlike for the other platforms—there exists a standardized and publicly available dataset (LFM-1b), which we enriched by crawling user connection information. Furthermore, the country distribution of the Last.fm dataset ensures a global scope. As a consequence, Table 3 presenting the results for the platform 500px exhibits four blank rows/columns because our 500px dataset does not contain users of Belarus, Czech Republic, Finland, and Mexico.

As the (per-row) normalized country connection matrix is asymmetric, it can also be used to address RQ1b, investigating which countries serve as an attractor for other countries. We define an *attractor measure* for a country c that models the (relative) amount of users from countries other than c that are attracted to establish connections with users in c . We define this attractor measure as the median of all shares of user connections from other countries maintained with users in c . This equals computing the median over all rows in the country connection matrix for the column representing country c (see, for instance, last row of the tables 2, 3, and 4). We use the median instead of the mean to correct for outliers, e.g., if only one or two countries account for a vast share of user connections to country c , we do not consider c as a *global* attractor.

To answer RQ1c concerning the role of the cultural dimensions by Hofstede for a country’s attractiveness, we compute Spearman’s rank order correlations between each country’s attractor value and each of the corresponding cultural dimensions.⁵

To answer RQ2, comparing the three OSN, we compute Pearson’s correlation coefficients between each country’s normalized country connection vector (i.e., respective row of the normalized country connection matrix) for each combination of the three OSN. The results are given in Table 5. Note that correlation coefficients could not be computed for the four countries that are not included in the 500px dataset.

⁵Since the ranges of Hofstede’s dimensions and attractor values highly diverge, we use rank order correlation.

Table 1. Number of users and user–user connections in the created datasets.

OSN	No. users	No. connections
Last.fm	55,191	1,087,662
500px	109,904	3,308,081
Facebook	106,249	166,129

3.3. Datasets

In the following, we describe in detail the data acquisition process from the three investigated OSN. Table 1 provides an overview of the number of users and user–user connections in the created datasets.

3.3.1. Dataset Last.fm. Last.fm is a social media platform for music enthusiasts. We use the publicly available LFM-1b dataset [21] including about 120,000 Last.fm users as starting point. Since we are interested in the cross-country relationships between users, we exclude all users for whom the dataset does not report country information, leaving us with about 46% of users. Subsequently, we use the Last.fm API endpoint *user.getFriends*⁶ to obtain all connected users. We cross-match these friends with the users in the LFM-1b, which eventually yields a total of 55,191 users and 1,087,662 user–user connections.

3.3.2. Dataset 500px. 500px is a social media platform for photography enthusiasts, amateurs and professionals alike. It is used as a photo sharing platform and facilitates direct user connections between photographers and clients.

As no suited dataset is publicly available, we had to create our own by implementing a crawler for 500px’s social graph. We use the 500px community REST API with several endpoints. First, we retrieve a set of random seed users—including user information such as country and number of friends—with the endpoint */v1/users*.⁷ Second, we use the endpoint */v1/users/:id/friends*⁸ to retrieve the respective users’ friend connections. The retrieved friends are added to a list and this list is used to fetch new friends. In other words, we employed a snowball system approach. Our final dataset contains 109,904 users with 3,308,081 user–user connections. Initially, the dataset resulted in 2,470 “countries” because the country information given by users often included city names and state

⁶<https://www.last.fm/api/show/user.getFriends>

⁷<https://api.500px.com/v1/users>

⁸<https://api.500px.com/v1/users/:id/friends>

names. To reduce noise in country information, we use Novatim⁹ to resolve the place of living into a country code and manually overviewed, extended, enhanced the country information in the dataset. Due to the uneven distribution over countries among users of 500px, this dataset does not cover all top 20 countries in the reference set (Last.fm). Therefore, Table 3 contains some empty rows and columns.

3.3.3. Dataset Facebook. In the absence of a suited publicly available dataset of Facebook user connections, we had to create our own by implementing a crawler for Facebook’s social graph. As Facebook does not provide an API for data access, we have to use the web interface and HTTP for communication and a HTML scraper for collecting the relevant data. For making web requests and parsing the response HTML, we used JSoup.¹⁰

We retrieve a user’s country information, total number of friends, and friend list (if publicly accessible). To fetch a user u ’s country information, we crawl u ’s “about” page and extract the specified “place of living”. We use Novatim¹¹ to resolve the place of living (which is typically a town or a country) into a country code. For each user in the publicly available friend lists (excluding duplicates), we repeat this procedure. To reach high dispersion, we selected random seed users from different continents to initiate the crawling process, and then follow a breadth-first-search strategy [22], which we adapted to obtain a roughly balanced amount of users per country. More precisely, we choose the source user u from a country that currently has the lowest number of crawled users. For that user u , we retrieve the friend list F_u , extract the country code of u ’s friends (if available), and record a connection between u and the elements in F_u . Our sample contains 106,249 Facebook users and 166,129 user-to-user connections.

While we are aware that this Facebook sample might have an uncharacteristic structure and may be affected by biases, and by bots, we believe that it can serve as starting point for a cautious comparative analysis between general-purpose and topic-specific OSN.

3.4. Data on Hofstede’s Cultural Dimensions

We use the base culture data (version 2015) for Hofstede’s six dimensions of culture as available from [23] and assign, for each country, the respective values of the dimensions to our dataset. Note that the values of Hofstede’s dimensions are in the range

⁹<http://nominatim.openstreetmap.org>

¹⁰<https://jsoup.org>

¹¹<http://nominatim.openstreetmap.org>

[1, 100].

4. Results

4.1. Cross-country Connections in the Country Connection Matrices

Table 2 shows the cross-country user connections, in relative numbers, for Last.fm, Table 3 for 500px, and Table 4 for Facebook. To answer RQ1a, we discuss all three tables.

In the OSN Last.fm (Table 2), in all countries, users are highly connected with users from the same country: all countries have values over 45% for within-country user connections, with the exceptions Canada (29.01%) and France (38.60%). The countries with the highest share of within-country connections are Poland (76.87%) and Brazil (76.04%) followed by Finland (65.27%), Czech Republic (64.21%), and Russia (62.80%). The strongest cross-country connections can be found for Canada with the United States (26.56%), Ukraine with Russia (19.22%), Belarus with Russia (17.12%), and Australia with the United States (16.47%). The relative numbers of other cross-country user connections are considerably lower, with most of them below 2%. For Canada, we find that the share of within-country user connections at about the same value as the share of connection of Canadian users with users from the United States (29.01% and 26.46%, respectively).

On 500px (Table 3), the strongest share of within-country user connections can be identified for the United States (26.89%), Germany (21.60%), and Russia (21.25%). The countries with the least share of within-country user connections are Sweden (9.67%) and Ukraine (11.13%). For some countries, the share of within-country user connections is lower than the share of connections with users from some other countries. For instance, within-country connections amount to 15.25% for Canada, but connections of Canadian users to US Americans to 18.25%. The same holds for Ukraine and Russia (within-country connections of Ukraine amount to 11.13%, while 16.41% of connections of Ukraine users go to Russia).

In general, if the share of user connections with another country is higher than the share of within-country user connections, then it is typically for cross-country connections with the United States, which is valid for Australia, Canada, Spain, the Netherlands, Poland, Sweden, and the United Kingdom. The share of connections from Poland to Russia (11.97%) is almost as high as the share of connections to users from the United States (12.40%). An exceptions is Ukraine, whose users connect to Russian users

(16.41%) more than to users from the United States (11.38%), which is again higher than the Ukrainian share of within-country user connections. Overall, the highest share of cross-country connections exists between Canadian users to users from the United States (18.25%), closely followed by connections from Australia to the United States (16.60%) and from Ukraine to Russia (16.41%).

Interestingly, the share of connections from Canadian users to users from the United States (18.25%) shows the fourth-highest share, preceded and followed by within-country shares only: United States (26.89%), Germany (21.60%), and Russia (21.25%), and on the fifth place France (17.60%).

On Facebook (Table 4), the countries with the highest within-country user connection shares are the Czech Republic (86.34%), Brazil (85.12%), and Mexico (81.50%). The other countries' within-country user connection shares are considerably lower, with shares lower than 1.50% within Australia, Spain, and Russia. For several country combinations, the share of cross-country connections exceeds that of within-country connections by far, particularly for those countries with a within-country user connection share of almost 0% (e.g., Australia and Spain). Particularly high shares of cross-country user connections maintain, for instance, the following country combinations: Canada–The Netherlands (58.36%), Japan–United States (46.25%), Ukraine–France (35.61%), Spain–United Kingdom, and Norway–Germany (34.93%).

Comparing the three OSN, it is interesting to see that the country pairs with the overall highest user connection shares on Last.fm and 500px typically share the same language (e.g., Australia–United States, Canada–United States, United Kingdom–United States) or have a shared second official language in their countries (e.g., Belarus–Russia, Ukraine–Russia)". On Facebook, in contrast, the highest shares of user connections are typically cross-language (e.g., Canada–The Netherlands, Japan–United States, Spain–United Kingdom, Ukraine–France).

The differences between the highest share of within-country user connections and the lowest one are far less pronounced in 500px compared to Last.fm, which is again far less pronounced compared to Facebook.

4.2. The Attractor Countries

The last row in the country connection matrices (Table 2 for Last.fm, Table 3 for 500px, and Table 4 for Facebook) presents the attractor measures (cf. Section 3.2) for the respective OSN and countries.

We use these to answer RQ1b.

In all OSN, the United States could be identified as strongest attractor country (7.07% for Last.fm, 13.44% for 500px, and 7.45% for Facebook). On Last.fm and Facebook, the measure ranges at about the same level (7.07% for Last.fm and 7.45% for Facebook). The distances to the respective next highest value are different, though, in the two OSN. On Last.fm, the United Kingdom could be identified as the second-place attractor country (4.22%). On Facebook, the distance to the second highest attractor value is larger with 3.09% for Germany. On 500px, the United States' attractor value is much higher (13.44%), followed by Germany (7.48%) and Russia (7.24%).

We further strive to identify whether the attractor measure (assuming that it is a proxy of a country's attractiveness) correlates with certain cultural aspects, in order to obtain clues on which cultural factors might make a country particularly appealing to users in other countries. An analysis with Spearman's rank correlation (cf. Section 3.2) identified, for the cross-country user connections, that users seem to be particularly attracted by countries that score high in Hofstede's cultural dimension *individualism*. For Facebook, the correlation is moderate ($\rho = 0.497$); for the other two OSN the correlation is weak to medium with $\rho = 0.341$ for 500px and $\rho = 0.301$ for Last.fm. Furthermore, the analysis also shows positive correlations for the dimension *masculinity*. The correlation is medium for 500px ($\rho = 0.409$) and Facebook ($\rho = 0.375$), but very weak for Last.fm ($\rho = 0.114$).

4.3. Differences Between the Online Social Networks

Comparing the country connection matrices (Table 2 for Last.fm, Table 3 for 500px, and Table 4 for Facebook) for the user connection patterns, we could identify OSN-specific particularities. On Facebook, there exists a large discrepancy between countries with very high within-country connections (> 50%) and countries with rather low (< 30%) within-country connections. On Last.fm, most countries have a high within-country connection ratio that is multiple times higher than the rather small ratios for cross-country connections. On 500px, the cross-country user connection shares are generally higher than on the other two platforms, while the shares of within-country user connections are rather low.

For answering RQ2, we delve into detail: Table 5 shows the correlations of the user connections between pairs of the three OSN, computed per country. For Last.fm and Facebook, the mean correlation over all countries is 0.649, indicating a medium to strong

correlation of user connection behavior on Last.fm and Facebook. For Last.fm and 500px, the mean correlation over all countries is 0.735, indicating a strong correlation of user connection behavior on Last.fm and 500px. For Facebook and 500px, the mean correlation over all countries is 0.485, indicating a weak to medium correlation of user connection behavior on Facebook and 500px. Correlating the results of *Last.fm and Facebook*, we find that for eight out of the top 20 countries, the correlation is > 0.9 , almost 1.0 for Brazil, the Czech Republic, Italy, and Mexico. Still, weak negative correlations indicate highly differing connection behavior between the two OSN for Australia, Spain, and Russia. Correlating the results of *Last.fm and 500px*, we find all correlations to be positive and basically strong (the correlation is > 0.6), except for Poland and Sweden with correlation coefficients of 0.439 and 0.489, respectively. The correlation is particularly high for the United States (0.962). Interestingly, the four countries that are not present in the 500px dataset, are countries with particular high correlation of Last.fm and Facebook user connection behavior. Correlating the results of *Facebook and 500px*, we find highly differing connection behavior between the two OSN for Spain and Russia. These countries, together with Sweden that has only a slightly positive correlation, are the ones with weak negative correlations for Last.fm and Facebook. High correlations can be found for Germany, Japan, and the United States. The other correlations are at a medium level.

Overall, our analysis reveals that the inclination to connect to users in the same country strongly differs between platforms. In fact, the average within-country connection share highly differs between OSN: Last.fm (54.45%), Facebook (39.79%), and 500px (15.38%). This indicates that the community of music enthusiasts (reflected on Last.fm) is much more likely to stay among their peers in the same country than the community of photographers (500px). A potential explanation is that music preferences are influenced by cultural background and market structures (e.g., local advertising campaigns, local radio airplay) and so people with similar interests are likely from the same country. The community of photographers may interact based on photo scenes or photography techniques, thus, on aspects that are not country-specific.

4.4. Addressing the Research Questions

Summing up, with respect to RQ1a (Which are the countries whose users have mainly intracountry user connections (i.e., connect with users from their own country)? Which are the ones that show transnational

connection behavior?) and partly addressing RQ2 (Are connection patterns (intracountry vs. transnational) comparable between different OSN?), the above described analysis shows the following: The connection patterns vary across OSN to a high degree. For instance, Brazil could be identified as a country characterized by a large share of intracountry connections on Last.fm and even more so on Facebook, while it ranges rather low in within-country user connections on 500px. Russia, with high values on the topic-specific OSN Last.fm and 500px, is the country with the least within-country user connection share on the general-purpose OSN Facebook.

Concerning RQ1b (What are the most important “attractor” countries, i.e., countries whose users are substantially more often the target of a friendship connection than other countries, in relative numbers?) and partly addressing RQ2 (Are connection patterns, i.e., intracountry vs. transnational, comparable between different OSN?), our results suggest that the United States can be considered a strong attractor country, surpassing other countries’ values. The level of attractor values, though, varies across OSN. Furthermore, different countries could be identified as second- and third-ranked attractor country on the different OSN (the United Kingdom and Russia for Last.fm; Germany and Russia for 500px; Germany and France on Facebook).

With respect to RQ1c (Is country attractiveness correlated with cultural aspects?), we found weak to moderate correlations between the attractor measures and cultural variables individualism and masculinity (in particular for Facebook and 500px).

Addressing RQ2 (Are connection patterns, i.e., intracountry vs. transnational, comparable between different OSN?), our results suggest that connection patterns vary across the analyzed OSN. On average, the connection patterns on Last.fm and 500px are more in line (mean correlation coefficient of 0.734) than each of these two OSN compared to Facebook. Still, our results do not allow to draw strong conclusions whether there are analogies in connection patterns between specialized OSN (Last.fm and 500px) versus general-purpose OSN (Facebook).

5. Practical Implications for Personalized Systems

Besides its contribution to the research area of social network analysis, we believe that the areas that will probably benefit most from our results are the fields of personalized systems and recommender systems. Several conclusions can be drawn from our analysis with regard to practical implications.

For instance, our results could help alleviating

Table 5. Pearson’s correlation coefficients between Last.fm (LFM) and Facebook (FB), Last.fm (LFM) and 500px, and 500px and Facebook (FB) connections per country.

Country	LFM-FB	LFM-500px	FB-500px
AU	-0.237	0.701	0.013
BR	0.996	0.654	0.642
BY	0.843		
CA	0.262	0.934	0.220
CZ	0.997		
DE	0.905	0.879	0.814
ES	-0.105	0.605	-0.156
FI	0.975		
FR	0.728	0.866	0.581
IT	0.992	0.747	0.679
JP	0.424	0.795	0.800
MX	0.992		
NL	0.925	0.659	0.578
NO	0.592	0.719	0.670
PL	0.771	0.439	0.569
RU	-0.065	0.888	-0.074
SE	0.885	0.489	0.429
UA	0.337	0.662	0.380
UK	0.848	0.754	0.718
US	0.915	0.962	0.897
Mean	0.649	0.734	0.485

the new-user cold start problem in recommender systems. Creating a user model based on the country-specific approximations of tie strengths and country’s attractor status would address this problem and could substantially decrease the new-user cold start problem. More concretely, today’s systems frequently use single sign-on buttons, which allow new users to register with their Facebook, Twitter, or other OSN accounts, giving the system access to their profile information. Since user profiles commonly contain country information, in the absence of item interaction data, our results could help trigger initial recommendations based on the typical connection patterns of users in the target user’s country.

Not only in cold start situations, also more generally, the information about cross-country user connections may be exploited to personalize recommendations depending on the target user and his or her connections to users in other countries. For instance, collaborative filtering (CF) techniques could be extended by a social tie strength filtering component, in a fashion similar to [24], where a CF recommender is adjusted to the target user’s preference for mainstream items by training on users with similar levels of “mainstreamness”.

Likewise, users with similar cross-country connection patterns could be clustered and served by a CF engine specifically trained on the cluster of the target user.

Besides item recommenders, also people recommender systems that suggest persons of interest to each other may integrate our findings into their algorithms. Depending on the target users’ needs to stay with others in the same country or to establish connections outside of their own country (e.g., if a student plans to go for a year abroad or is on currently abroad), recommenders could adjust the distribution of recommended people inside and outside of his or her country accordingly.

6. Conclusion and Outlook

In this work, we examined the nature of personal networks and user connections in OSN across countries for three different OSN.

Our contributions may be summarized as follows: First, while, theoretically, OSN enable users to transcend geographical borders and OSN indeed show global interconnections on a wide basis, our analysis identified that people in some countries are largely interconnected with people from their own country. In other words, the generally high shares of within-country user connections show that the analyzed OSN are far from representing a “global village”. At the same time, our analysis shows that cross-country connections interconnect all the analyzed countries to a certain degree. Our Facebook dataset, though, shows some exceptions (e.g., Canada, Russia, Spain, Sweden).

Second, overall our analysis suggests that both, country-specific connection patterns as well as platform-specific patterns, are represented in OSN. In other words, the inclination to connect to users within the same country strongly differs between platforms. We argue that not only the individuals’ country-specific but also topic-related sub-communities have to be considered when studying user connections in OSN.

Third, our analysis identified “attractor” countries, whose users seem to be particularly attractive for user connections from other countries. We found that countries scoring high in individualism and masculinity are particularly attractive for user connections.

As one research avenue in the context of the work at hand, we contemplate a detailed comparison of the levels of cross-country user connections and country-specific and socio-economic indicators, similarly to [25], where we correlate country similarities in terms of their citizens’ music taste and indicators such as Hofstede’s cultural dimensions e.g., masculinity or indulgence) [15] and socio-economic factors according

to the Quality of Government dataset¹² (e.g., GDP, income inequality, or unemployment rate).

An interesting problem for future research that contributes to the field of recommender systems is to investigate whether recommender systems perform better when a social tie filtering component is based on the user connection patterns of a respective topic-specific OSN (for instance, using the Last.fm user connection patterns for a music recommender system) or when it is based on the user connection patterns of a general-purpose OSN (e.g., Facebook).

Acknowledgments

This work is supported by the Austrian Science Fund (FWF): V579. The authors would further like to thank Dominik Bryant for developing the 500px crawler and Markus Hirth, Markus Laube, and Martin Peché for developing the Facebook crawler.

References

- [1] M. McLuhan, *The Gutenberg Galaxy*. Toronto, Canada: University of Toronto Press, 1962.
- [2] M. McLuhan, *Understanding Media: The Extensions of Man*. Toronto, Canada: MacGraw-Hill, 1964.
- [3] N. B. Ellison *et al.*, “Social network sites: Definition, history, and scholarship,” *Journal of computer-mediated Communication*, vol. 13, no. 1, pp. 210–230, 2007.
- [4] D. Boyd, “Why youth (heart) social network sites: The role of networked publics in teenage social life,” *MacArthur foundation series on digital learning—Youth, identity, and digital media volume*, pp. 119–142, 2007.
- [5] R. Grieve, M. Indian, K. Witteveen, G. A. Tolan, and J. Marrington, “Face-to-face or facebook: Can social connectedness be derived online?,” *Computers in Human Behavior*, vol. 29, no. 3, pp. 604–609, 2013.
- [6] Y. Kim, D. Sohn, and S. M. Choi, “Cultural difference in motivations for using social network sites: A comparative study of american and korean college students,” *Computers in human behavior*, vol. 27, no. 1, pp. 365–372, 2011.
- [7] N. Nash, “International facebook “friends”: Toward mcluhan’s global village,” *The McMaster Journal of Communication*, vol. 5, 2010.
- [8] M. Bailey, R. R. Cao, T. Kuchler, J. Stroebel, and A. Wong, “Measuring social connectedness,” tech. rep., National Bureau of Economic Research, 2017.
- [9] G. A. Barnett and G. A. Benefield, “Predicting international facebook ties through cultural homophily and other factors,” *New Media & Society*, vol. 19, no. 2, pp. 217–239, 2017.
- [10] D. D. Beer, “Social network (ing) sites revisiting the story so far: A response to danah boyd & nicole ellison,” *Journal of Computer-Mediated Communication*, vol. 13, no. 2, pp. 516–529, 2008.
- [11] S. M. Choi, Y. Kim, Y. Sung, and D. Sohn, “Bridging or bonding? a cross-cultural study of social relationships in social networking sites,” *Information, Communication & Society*, vol. 14, no. 1, pp. 107–129, 2011.
- [12] S. A. Myers, A. Sharma, P. Gupta, and J. Lin, “Information network or social network?: the structure of the twitter follow graph,” in *Proceedings of the WWW 2014*, pp. 493–498, ACM, 2014.
- [13] H. Kwak, C. Lee, H. Park, and S. Moon, “What is twitter, a social network or a news media?,” in *Proceedings of the 19th International Conference on World Wide Web, WWW ’10*, (New York, NY), pp. 591–600, ACM, 2010.
- [14] M. Chan, “Social network sites and political engagement: Exploring the impact of facebook connections and uses on political protest and participation,” *Mass Communication and Society*, vol. 19, no. 4, pp. 430–451, 2016.
- [15] G. Hofstede, *Cultures consequences: comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: Sage publications, 2001.
- [16] G. Hofstede, G. J. Hofstede, and M. Minkov, *Cultures and Organizations: Software of the Mind*. New York, NY: McGraw-Hill, 3rd ed., 2010.
- [17] DMR, “How many people use facebook, youtube and other social media.” <https://expandedramblings.com/index.php/resource-how-many-people-use-the-top-social-media>, February 2018. [accessed 2018-9-3].
- [18] J. J. Jones, J. E. Settle, R. M. Bond, C. J. Fariss, C. Marlow, and J. H. Fowler, “Inferring tie strength from online directed behavior,” *PloS one*, vol. 8, no. 1, 2013.
- [19] E. Gilbert and K. Karahalios, “Predicting tie strength with social media,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, (New York, NY), pp. 211–220, ACM, 2009.
- [20] K. Hampton, L. S. Goulet, L. Rainie, and K. Purcell, “Social networking sites and our lives,” *Pew Internet & American Life Project*, vol. 16, pp. 1–85, 2011.
- [21] M. Schedl, “The LFM-1b Dataset for Music Retrieval and Recommendation,” in *Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR)*, (New York, NY), June 2016.
- [22] S. A. Catanese, P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, “Crawling facebook for social network analysis purposes,” in *Proceedings of the International Conference on Web Intelligence, Mining and Semantics, WIMS ’11*, (New York, NY), pp. 52:1–52:8, ACM, 2011.
- [23] Hofstede *et al.*, “Dimension data matrix.” <http://www.geerthofstede.nl>, 2015. [accessed 2018-9-3].
- [24] M. Schedl and C. Bauer, “Introducing Global and Regional Mainstreamness for Improving Personalized Music Recommendation,” in *Proceedings of the 15th International Conference on Advances in Mobile Computing & Multimedia (MoMM 2017)*, (Salzburg, Austria), December 2017.
- [25] M. Schedl, F. Lemmerich, B. Ferwerda, M. Skowron, and P. Knees, “Indicators of Country Similarity in Terms of Music Taste, Cultural, and Socio-economic Factors,” in *Proceedings of the 19th IEEE International Symposium on Multimedia (ISM 2017)*, (Taichung, Taiwan), December 2017.

¹²<https://qog.pol.gu.se/data/datadownloads/qogoeccdata>