

Madrid, Spain 2 November, 2011

Twitter: Who follows whom?

[Twitter](#) is currently one of the most popular and fastest growing applications in the internet. Created in 2006 by Jack Dorsey and Biz Stone, it is a micro-blogging system which has rapidly attracted a large number of users to become one of the most successful platforms for both social interaction and information diffusion. On Twitter, a user can post messages, otherwise known as "tweets", of up to 140 characters. Twitter is currently used by around 200 million internauts who upload more than 140 million tweets to the system every day.



Ruben Cuevas [Universidad Carlos III de Madrid](#) and Balaji Rengarajan [Institute IMDEA Networks](#)

While principally a social network, Twitter has also turned into a disruptive communication medium which is being used for various purposes. Apart from its obvious use as a conversational tool, it has been used as a viral marketing tool, and even to organize protests. Twitter played a key role in the Arab Spring, which involved the 2011 Egyptian revolution, as well as the 2009/2010 Iranian election protests by enabling protesters to communicate with each other and to organize themselves. Twitter also made it possible for news from these regions to reach the outside world in real-time, despite the large-scale blackouts of the traditional media.

The main foundation on which Twitter is based is its ability to allow one user to follow others that are of interest to them. Any given user registered on the system can follow any other user on the system, e.g. Alice. We then refer to Bob as one of Alice's followers, and Alice as Bob's friend. This friend-to-follower relationship (or link) allows Bob to see every tweet posted by Alice. Each user will see the mixture of tweets generated by other users who he/she follows in reverse chronological order. Tweets can potentially be re-tweeted by a user to their followers, generating tweet chains.

A research study undertaken by the [NETCOM](#) research group at [University Carlos III of Madrid](#), an active collaborator with [Institute IMDEA Networks](#), has provided some insight into who uses Twitter and who they follow, enabling us to gain some understanding on the social impact of this social network. The study focus on both the locations of users and their followers to determine if they are geographically concentrated and, if so, to identify where. In this article we shall attempt to answer the following question: Are followers typically located close to their friends? Understanding such a Locality phenomenon of large-scale systems such as Online Social Networks (OSNs) is critical in order to improve system design and users' performance, while reducing infrastructural and operational costs. To capture Locality effects, the metric the study uses is user-level distance, which captures a representative metric per user, such as the median distance to its followers. To conduct the study, researchers collected real data including the geographical location of around 1 million Twitter users, and more than 16 million of their followers. Overall, the dataset includes more than 100 million friend-to-follower links.

Researchers observed that about 40% of the links have an associated distance lower than 1,000 km. This represents intra-country communications for most countries in the dataset. Furthermore, they observed that 80% of the links are in a range of 4,000 km, which involve intra-country communications for large countries such as the USA or Brazil, and intra-continent relationships for Western Europe. However, there are still around 10% of long-distance links over 6,500 km that represent cross-continent links. Therefore, while most users have followers that are relatively proximate, there are also a substantial number of long-distance relationships. What's more, popular users (i.e. those with a larger number of followers) are responsible for most of the long-distance links and the typical distance to their followers is larger than that of unpopular users. Thus, non-local relationships do play a key role in the Twitter network. However, this global analysis is clearly influenced by the dominance of the USA, which represents 50% of the friends, followers and links. As such, the researchers have deepened and broadened the study by analyzing the influence of geo-political, cultural and language factors on the Twitter ecosystem.

Thus, the study groups the users by country, and examines the relationship between users in different countries and their respective followers. Researchers have selected the users' country as a criterion, since it allows them to accurately group together those users who have a proximate geographical location, a similar cultural profile and the same language. Furthermore, they have selected the 15 countries that contribute the largest number of friends to the dataset. From the language perspective, the study differentiates countries into two profiles. On one hand, we have those countries whose official (or co-official) language is English, such as the USA, the UK, Canada, Ireland, India and Australia. On the other hand, we have those countries with an official language other than English, such as Brazil, Spain, Germany, France, Italy, Indonesia, Japan and the Netherlands. As expected, the observed global locality trends do not apply to every country. Based on the study's observations we can distinguish 3 different profiles:

Local profile: This is composed of a group of countries whereby an overwhelming majority of

users have followers that are mostly in the same country. This profile includes the USA on one hand, and countries such as the Netherlands, Brazil, Indonesia, Germany and Spain on the other, i.e. those which have an official language other than English. We observed that around 90% of users from the USA typically have a distance to their followers which is less than 4,000 km, which defines the boundary of intra-country relationships for the USA. Firstly, this is due to the prominence of users from the USA on Twitter, and secondly to their strong local culture. This intra-country locality effect is even more impressive in the case of Brazil, where 90% of the users have a user level distance of less than 2,000 km, where the limit of intra-country relationships is also about 4,000 km. This confirms the presence of regional-based locality in Brazil. While the other aforementioned countries demonstrate lower locality, the majority of the users in these countries have followers who are local.

Shared Locality profile: This category consists of those countries which contain a roughly equal number of users having mainly local followers, and users who have many followers outside the country. Examples of such countries include France, Mexico, Italy and Japan, where Twitter enjoys comparatively lower popularity. In France, 60% of users have a median distance to their followers that is lower than 1,000 km. However, several neighboring countries such as the Netherlands, Belgium, Switzerland, Italy and Germany are located within this range. As such, a portion of this 60% represents inter-country relationships rather than intra-country ones. Finally, around one third of French users have a typical distance to their followers of between 5,500 and 9,500 km, which represents the population of followers in the USA.

English-based (external) Locality profile: This profile is composed of countries where English is the official or co-official language. They experience an important external Locality with many users having a substantial number of followers in the USA. Furthermore, an important portion of users have mainly local followers. If we analyze users from the UK, we can see the bi-polarity described above between the UK and USA. Fifty percent of users have mainly local followers within a range of 1,000 km, while 37% of users have followers who are predominantly in the USA.

Another interesting observation is that user popularity plays a different role in different countries. With regard to users in the USA, the median distance to followers increases steadily with increasing user popularity. However, in the case of Brazil, popularity does not seem to affect the median distance of followers at all. In the case of countries such as the UK and France, there is a clear division among users, with popular users having a significant number of followers that are in the US, while the more obscure users have predominantly local followers.

Understanding such underlying patterns of relationships in social networks allow us to obtain insight into the social impact of these networks through the unprecedented volume and range of information transfer which they enable. It also helps us to better understand how cultural and societal characteristics mould social networks. Understanding the Locality effect of internet scale systems also has direct implications on improving the performance of such systems and building better next-generation systems. Social networks are a very powerful tool for human interaction which demands increasing attention by the international scientific community. The development of the Internet cannot be properly understood, monitored and optimized without a close understanding of the dizzying evolution of online social networking.

Read more:

- [Madrimsd - notiweb](#) (in Spanish)

Source(s): IMDEA Networks Institute

URL: [Twitter: Who follows whom?](#)

About us

IMDEA Networks Institute, promoted by the Regional Government of Madrid, is a research organization on computer and communication networks whose multinational team is engaged in cutting-edge fundamental science and technology. As a growing, English speaking institute located in Madrid, Spain, IMDEA Networks offers a unique opportunity for pioneering scientists to develop their ideas. IMDEA Networks has established itself internationally at the forefront in the **development of future network principles and technologies**. Our team of highly-reputed researchers is designing and creating today the networks of tomorrow.

Some keywords that define us: 5G, Big Data, blockchains and distributed ledgers, cloud computing, content delivery networks, data analytics, energy-efficient networks, fog and edge computing, indoor positioning, Internet of Things (IoT), machine learning, millimeter-wave communication, mobile computing, network economics, network measurements, network security, networked systems, network protocols and algorithms, network virtualization (software defined networks – SDN and network function virtualization – NFV), privacy, social networks, underwater networks, vehicular networks, wireless networks and more...

IMDEA Networks Institute
28918 Leganes (Madrid) Spain
Avda. del Mar Mediterráneo, 22

+34 91 481 6210
mediarelations.networks@imdea.org
www.networks.imdea.org

Twitter: [@IMDEA_Networks](#) | [Facebook](#) | [Instagram](#) | [Flickr](#) | [YouTube](#)