NETS 150 Final Project

Introduction

In today's fast paced and connected world, breaking news and information travels across the world in a matter of seconds. With world reaching news channels and global websites, news from any part of the Earth, no matter how remote, is a few clicks away. However, interest from users differs in different countries, dependent on the applicability and the exposure given to articles pertaining to a news topic, and further dependent on the access and familiarity with modern technology and common language. We were interested in documenting and studying the diffusion of news pieces across countries. To limit the effects of the above mentioned difficulties, we searched for a topic that unites the world, transcends all geographic, social and economic boundaries. We ended up deciding on sports, specifically the game of football (soccer). All around the world, there are thousands of soccer fans, making it the most popular sport and the FIFA World Cup the most watched sporting event on the planet.

Thus using the power of the Internet, we plan to document and track the dissemination of the most recent football news into different countries to act as an estimation measure of the effectiveness as the Internet as a tool for organic spread of information with people acting as vertices in the connected graph. By measuring the spread we can discern better connected countries to the rest of the world.

Hypothesis

Due to the longstanding love and popularity of the sport of football in Europe and South America, we expect to see those countries to show a faster spread and greater discussion about breaking news. By setting the Guardian and the UK as the center, which has one of the highest readerships and most number of international soccer players respectively, we chose an information hub that everyone had an opinion about. By examining the public discourse through the proxy of social networks, we can measure and rank the countries' spread of football news across the Internet.

To best represent a global picture we chose 8 countries from across the world spanning 6 continents: Australia, Brazil, Canada, France, India, Netherlands, South Africa, UK, USA.

From previous studies of Facebook posts during the World Cup, we made an assumption of each our countries ranking, which is documented in our results.

Global World Cup Conversation on Facebook

Top Quarter

Bottom Quarter

As a percentage of all people who are active on Facebook every day.
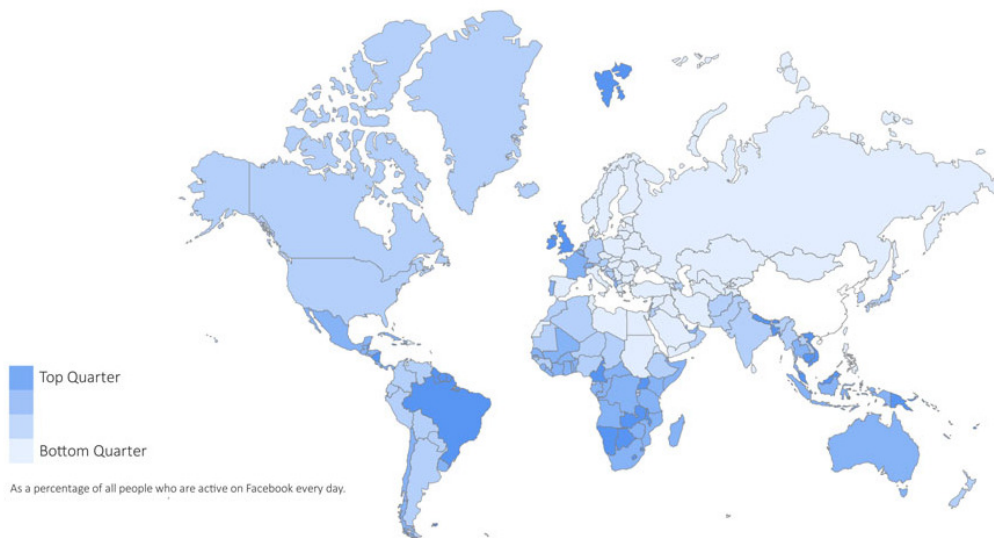
Method

We studied this effect by looking at popularity and significant discussion of football news articles on social networks, specifically Twitter. Twitter allows us to access tweets in a given city in the past two weeks. To estimate the most talked about topics of discussion in football, we searched the tweets from the three most populated/important cities of the country, in the past two weeks that were about football. The Guardian was used to provide articles and details on all the events and breaking news of the same time period in the Football Section. All the tweets from a given country were compiled into an individual text file, and all the text files together with the articles served as the corpus in the vector space model. The tweets files were compared to each and every article. By averaging the similarity score, we ranked the countries and thus could estimate the overall spread of the sports stories.

The reason that accumulating all the tweets into a single file made more sense for us because the number of tweets and corresponding documents for each tweet would become too large to compare properly.

Results

COUNTRY	AVERAGE SIMILARITY SCORE
AUSTRALIA	30.45302
BRAZIL	10.81636
CANADA	16.76076
FRANCE	9.131747
INDIA	29.40119
NETHERLANDS	8.478055
SOUTH AFRICA	59.96801
UK	45.01611
USA	36.13139

Popularity by Country
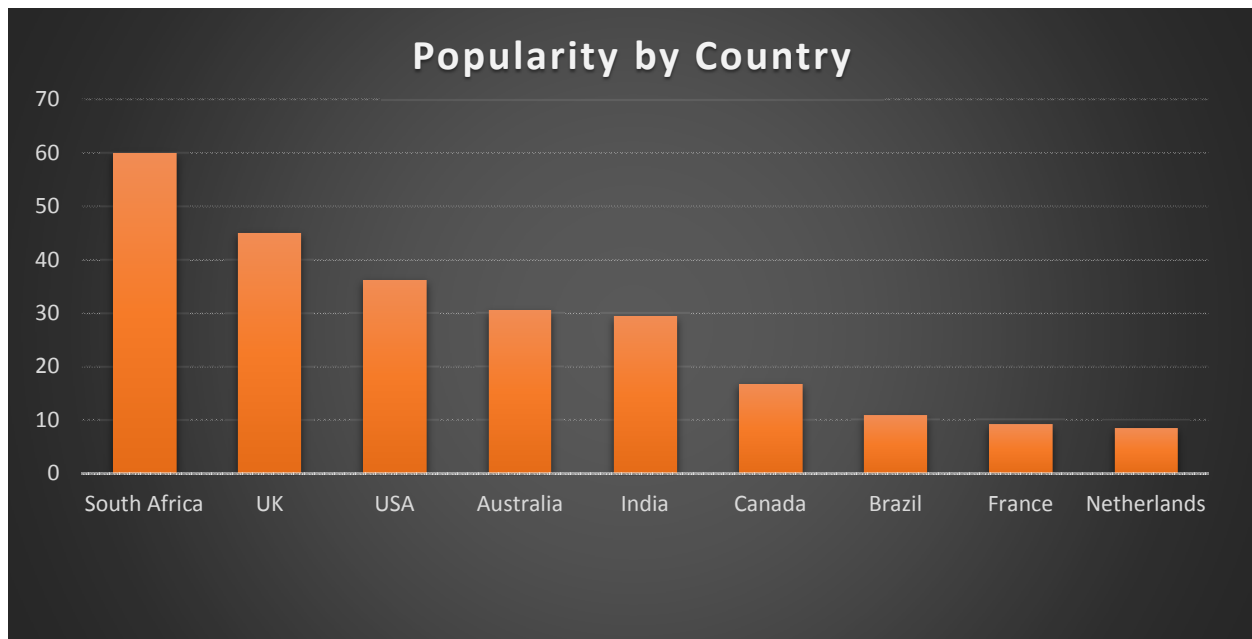
Analysis

One of the most surprising results of our analysis was that South Africa had the highest similarity score. A cursory search on the Twitter website leads to speculation that this may have occurred due to the presence of large number of tweets from the account of SuperSports, a sports channel that tweets about every major football match and breaking news on the topic.

It comes as no surprise that the news center of our Analysis model, the United Kingdom came second. The ranking of popularity in other countries seems to follow a reasonable pattern as well. India's high ranking can be attributed to the large number of English speakers in the country and the popularity of the English Premier League.

The high ranking of the United States of America points to a growing and vocal fan base in a country that has historically never been interested in the game. They also have the largest number of twitter users.

While football heavyweights Brazil, France and Netherlands make up the bottom three of the ranking, this may be due the wide spread usage of Portuguese, French and Dutch, something that our model could not account for because the Guardian is published in English.

RANK	EXPECTED RANKHOLDER	ACTUAL RANKHOLDER
1	UK	South Africa
2	Brazil	UK
3	France	USA
4	South Africa	Australia
5	Australia	India
6	USA	Canada
7	India	Brazil
8	Netherlands	France
9	Canada	Netherlands