Web Social Network (ORKUT)

Suhas Phand (05CS6013)
Sandeep Karajgaonkar (05CS6018)

M.Tech(CS), 1st year,
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur

1 Abstract

Now a days online social networks are increasing rapidly, Orkut is one of them. In this project we will study various properties of orkut network. As orkut is mainly friendship network, as friendship is mostly considered as transitive relation, we will see if it is actully reflected in the network by determining its clustering coefficient. Here we will find out what common properties the friends do hold, like age, country, religion, smoking or driking habits etc. We will also compare the networks of indian, brazilian and pakistani people to find if there is any difference in the properties of network.

2 Introduction

A social network is a map of the relationships between individuals, indicating the ways in which they are connected through various familiarities such as age, city, college, company, and other interests. Now a day such social network is also growing on internet such as Orkut which helps people to interact with other peoples of network. This social network is still invite only i.e. any one cant join the network unless he/she is get invited by the current member. Orkut has released by Google named after Orkut Buyukkokten, a Google software engineer who developed the project during personal time allowed by Google, which allowed the professional to spend twenty percent time working on personnel interest. In this we are going not analyse this Orkut network through social properties to find out similarities between the nodes and other interesting facts.

3   User Registration and Data

As mentioned above Orkut is still invite only so new user cant join the Orkut unless he is invited by the current member and do registration. While registering, users entered their names, e-mail addresses, birthdays (for birthday reminder notifications to their friends), educational details (graduate or undergraduate status) and year in school, residence, and home country and state. As well as their phone number, hometown, homepage and picture , their interests to join this network such as friendship, networking, activity partner , dating etc . In addition to basic demographic information users were asked to add a list of interests and hobbies to their profile by checking off as many choices as they liked from listings of interested activities such as sports, movie, music, and books. Many of the information is optional one and user can restrict this to friends or friends of friends or allow to anyone to view that. While analyzing we are assigning the unique IDs to each node i.e. profile. The resulting dataset was a social network with profile for each of the members. In next sections we analyze the Orkut from a network perspective and then go into detail and look at relationship between the user attributes.

4   Network Analysis

Orkut is huge online network more than 16 million peoples using Orkut. So we are going to analyse the part of network belonging to Brazil as majority of users around 71.94% user belongs to it and India (3.06%) and Pakistan(2.08) . individually and analyze them on the following -

- Degree distribution

- Age distribution

- Clustering coefficient

- Distance distribution (Small world effect)

- Male to female ratio

4.1   Degree distribution

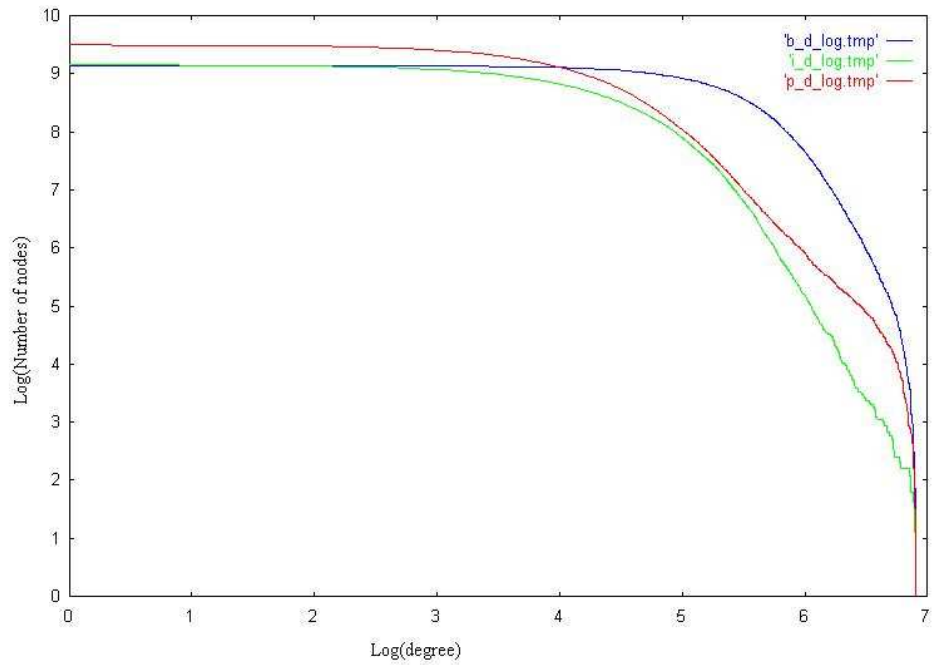Following graph shows the degree distribution of the respective network.

Figure 1: Degree distribution of the respective network

4.2 Age distribution

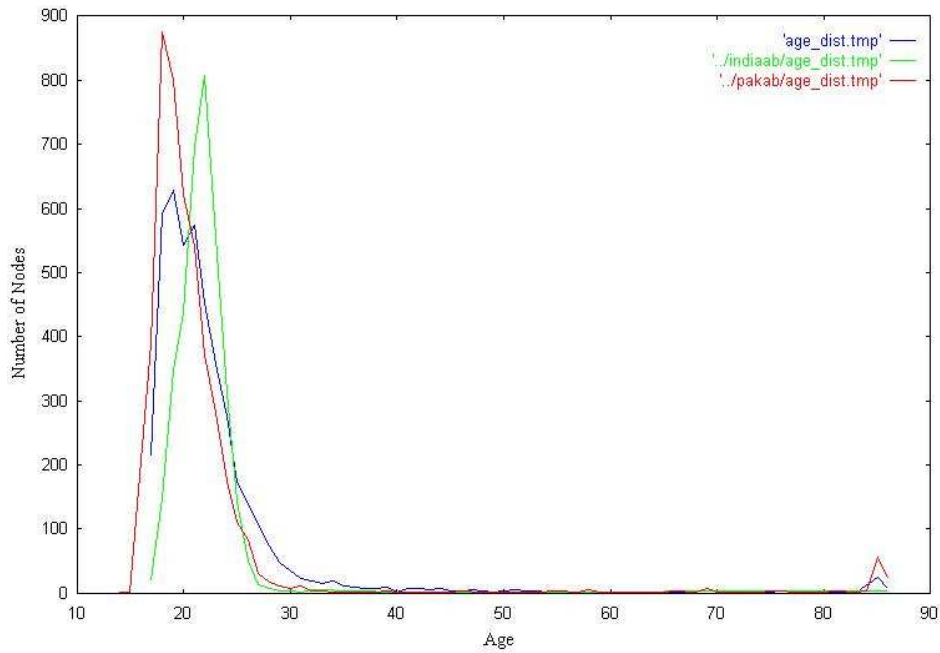Following graph shows age distribution of the network.

Figure 2: Age distribution of the respective network

Most of the users are belongs to age group of 19-23.

4.3 Clustering coefficient

It is a measure of transitivity which us how many of a user's friends, friends of themselves are. Following table gives the average clustering coefficient of the Brazil, India and Pakistan network.

Brazil	India	Pakistan
0.201379	0.239391	0.236432

4.4 Distance distribution

	Averagedistanceamongpair	Diameter
Brazil	5.46716	10
India	4.43172	9
Pakistan	5.78772	11

Thus all three networks show the small world effect. But question still remains whether entire network shows the small world effect or not? But Orkut has shown small world effect by [2] which studied the Orkut by late 2004.

4.5 Male female ratio

	Male%	Female%
Brazil	42.7638	57.2362
India	79.1069	20.8931
Pakistan	62.277	37.723

4.6 Here for

	ActivityPartner%	Friends%	Dating%	BusinessNetworking%
Brazil	15.4	60.2	05.4	18.96
India	19.60	52.58	11.55	16.26
Pakistan	18.71	63.03	06.99	11.27

While the above analysis of the network topology is insightful, things become even more interesting when user profiles are taken into account. We will explore these profile features in the next section and will later return to their impact on network properties.

5   Matching profile with friend

In the process of registering users were asked describing their personalities, for what they are here such as friendship, networking event partners, dating etcHere we are trying to study a node and its neighbors i.e. friends attributes or such as age, language, here for and other views to find out whether these things plays a role in making a friend or not.

As we are studying three different parts of network we are going to analyze these properties in three of them.

Age-

Here we have found that there are large no. of neighbors i.e. friends having same age is large.

$Age difference$	$Number of friends\%$
0	34.65
1	78.43
2	92 .72
3	97.52
4	99.13

From above we can say that in friends any body will have high no of the friends of same age than other age.

5.1   Relationship status-

Here we analyze the relationship status of neighbors.

	$Brazil$	$India$	$Pakistan$
$single$	0.515620	0.941644	0.844280
$committed$	0.466255	0.074521	0.164655
$married$	0.197822	0.031915	0.144943

This data shows that in Brazilian network all status people have joined the network where as in India and Pakistan it all about singles.

5.2 Smoking habit

	Brazil	*India*	*Pakistan*
No	0.913274	0.896121	0.804006
Yes	0.151697	0.167620	0.261436

5.3 Drinking habit

	Brazil	*India*	*Pakistan*
No	0.299533	0.783042	0.876391
Yes	0.782573	0.284642	0.202743

5.4 Friends gender

	Brazil	*India*	*Pakistan*
Female	0.359073	0.571444	0.528365
Male	0.856845	0.428007	0.736750

Following table shows the average values that your friends follow your attributes.

	Brazil	*India*	*Pakistan*
Age	0.757801	0.897633	0.748770
Country	0.733555	0.993953	0.842383
Religion	0.363388	0.808404	0.940526
Relationship status	0.465577	0.890078	0.726510
Sex	0.511845	0.765405	0.660830
Smoking habit	0.847285	0.817014	0.689472
Drinking habit	0.667985	0.669492	0.787170
Degree	0.048771	0.119241	0.115365
Clustering coefficient	0.201379	0.239391	0.236432
Living	0.093950	0.140356	0.147009
Language	0.471640	0.925763	0.907019
Humour	0.041235	0.143898	0.088168
Here for	0.465022	0.942842	0.724036
Political View	0.014225	0.121712	0.125259

6   FACTS

Following figures shows the actual information about overall Orkut network. This information is taken from www.orkut.com and is getting updated daily.
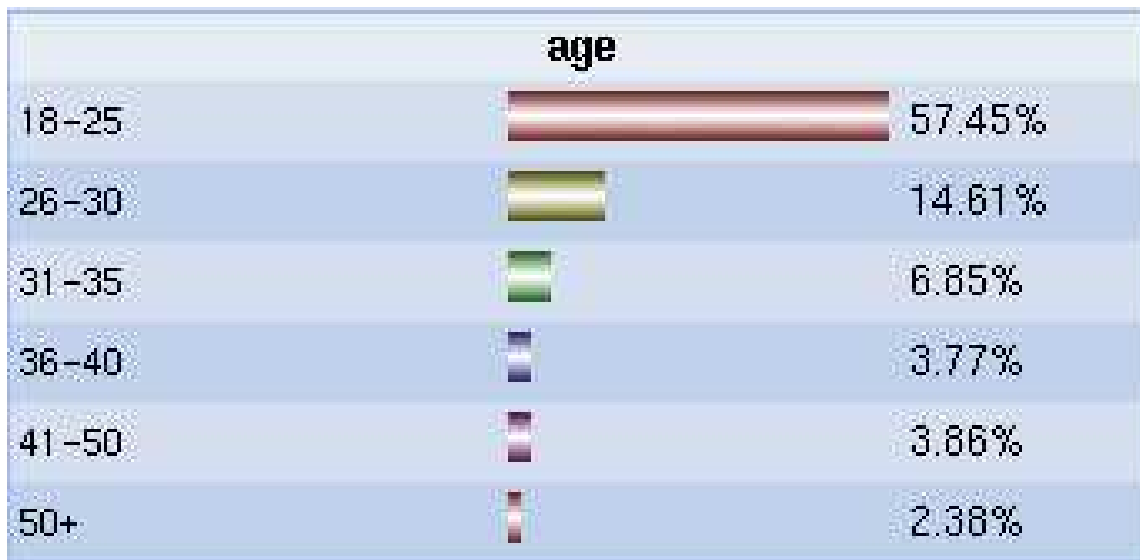
Age

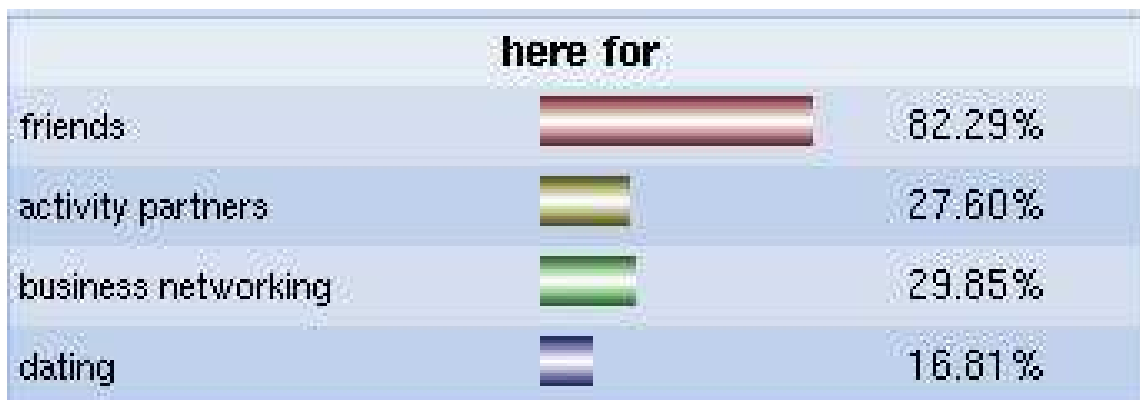Figure 3: Actual age distribution in ORKUT network

Here for

Figure 4: Actual here for distribution in ORKUT network

Relationship status

Figure 5: Relationship status in ORKUT network
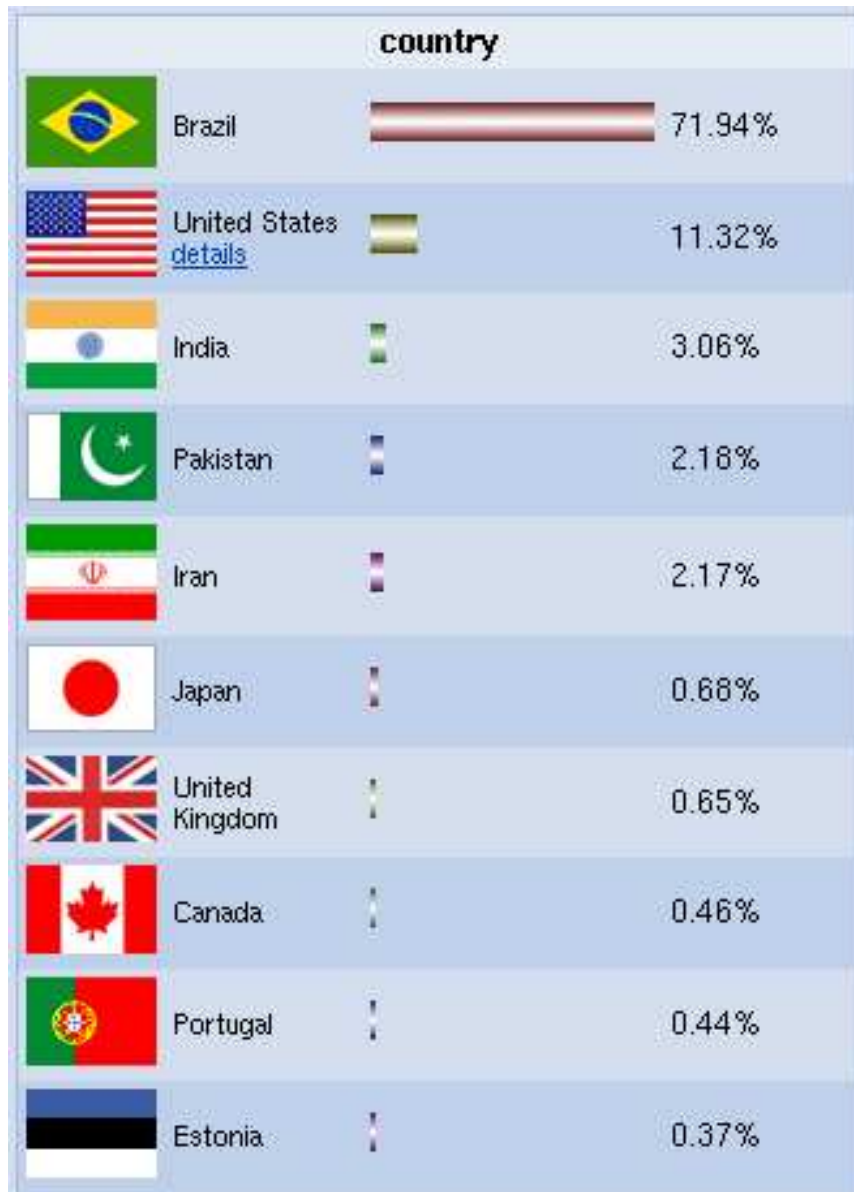
Country

Figure 6: Countries in ORKUT network

7   Conclusion

In Orkut network, people of same age, from same country and speaking same language comes together. No other attributes really matter. Age group 19-23 dominates the network and as provided by Orkut Brazilians dominates the network.

The network shows low clustering coefficient (0.2), which means transitivity is not followed in the network as expected in friends.

Comparing Brazilian, Indian and Pakistani network, main difference in these networks is male-female ratio. In Brazilian network more than 50% are female, while for India and Pakistan this value is very small.

References

[1] www.orkut.com

[2] Zahid Anwar William Yurcik Vivek Pandey, Asim Shankar, Indranil Gupta, Roy H. Campbell**Leveraging Social-Network Infrastructure to Improve Peer-to-Peer Overlay Performance: Results from Orkut**