

Social Network Analysis of the Caste-Based Reservation System in India

Akrati Saxena¹, Jaspal Singh Saini¹, Yayati Gupta¹, Aishwarya Parasuram²,
Neeharika³, and S.R.S. Iyengar¹

¹ Department of Computer Science and Engineering, Indian Institute of Technology
Ropar, India

{akrati.saxena, jaspal.singh, yayati.gupta, sudarshan}@iitrpr.ac.in

² Google Alphabet

aparasuram@google.com

³ Department of CSE, India

neeharika@gmail.com

Corresponding Author: S. R. S. Iyengar

Abstract. It has been argued that the reservation system in India, which has existed since the time of Indian Independence (1947), has caused more havoc and degradation than progress. This being a popular public opinion, has not been based on any rigorous scientific study or research. In this paper, we revisit the cultural divide among the Indian population from a purely social network based approach. We study the distinct cluster formation that takes place in the Indian community and find that this is largely due to the effect of caste-based homophily. To study the impact of the reservation system, we define a new parameter called social distance that represents the social capital associated with each individual in the backward class. We study the changes that take place with regard to the average social distance of a cluster when a new link is established between the clusters which in its essence, is what the reservation system is accomplishing. Our extensive study calls for the change in the mindset of people in India. Although the animosity towards the reservation system could be rooted due to historical influence, hero worship and herd mentality, our results make it clear that the system has had a considerable impact on the country's overall development by bridging the gap between the conflicting social groups. The results also have been verified using the survey and are discussed in the paper.

1 Introduction

The Caste System in India

In the context of the Indian society, caste is defined to be a Hindu hereditary class of socially equal persons, united in religion and usually following similar occupations, distinguished from other castes in the hierarchy by its relative degree of spiritual purity or pollution [1]. It is said that the origin of the caste system is credited to the *Vedas*, the mythological texts that claim the very basis of Hindu religion, according to which, the primal man destroyed himself to

create a human society. The different castes were created from different parts of his body. The *Brahmins* (scholar or priest class) were created from his head, the *Kshatriyas* (soldier class) from his arms, the *Vaishyas* (business class) from his thighs, and the *Shudras* (menial labor class) from his feet. The hierarchy in the caste is determined by the descending order of importance of his body organs.

Another religious theory claims that the caste system was created from the body organs of Brahma, who is believed to be the creator of the world according to the Hindu religion. This stratification, though an obvious myth, has stayed in the Indian society since time immemorial.

Economic Impact of the Caste System

Although the caste groups were supposedly divided along the lines of spiritual purity, they soon came to determine inflexible occupational roles. The downside of this system surfaced with the birth of caste-based discrimination, which is still a dominant phenomenon today. As time progressed, one's caste became intrinsically linked with one's wealth, social status, and even entry into public places. The society became largely dominated by the so called upper castes, and the rest were denied economic freedom, forced to work menial jobs and prevented from trying to improve their economic status. This led to the concentration of assets, social capital, and power in the hands of one section of the society. Therefore, while the socially forward classes progressed, the socially backward classes lagged behind in terms of literacy rates, education levels, income levels, and other measures of socio-economic well-being.

The Reservation System in India from a Network Theory Perspective

To balance this routine of prejudicial social stratification, affirmative steps were undertaken to uplift the backward classes, and the reservation system was introduced wherein a certain number of seats are reserved for the members of socially and economically backward classes at the places of higher education and government jobs. However, it was soon met with a lot of backlash from the socially forward community, who felt that this system is not meritorious, and provides an undue advantage to members of the socially and economically backward classes.

In this paper, we study the existing reservation system from a pure network-theoretic perspective. The network science concepts like homophily [2], weak ties [3], social distance, opinion formation [4], influence propagation [5,6,7], can help to explain the positive outcomes of the reservation system. Here, we consider two social communities, the *socially forward and uplifted* (FC) and the *socially backward and downtrodden* (BC). We establish why and how the reservation system maintains a very good balance between the two.

Past studies in network analytics by Jackson has shown that network formation and subsequent interaction between the nodes is highly influenced by homophily [8]. In India, associations amongst the people are seen to be largely determined by caste-based homophily, hence, for the purposes of our study, we

choose to term the network formation pattern among the Indian population as a *caste-based homophilic network*.

Our motivation comes from the existence of a tangible strength associated with every weak tie, as proposed by Granovetter in his famously cited theory of the *Strength of Weak Ties* [3]. He observed the importance of weak ties to get new opportunities. This is a very prominent network phenomenon that has been ignored in past studies of the reservation system, which we have chosen as the key element of our study. We assume whenever a reservation is given, it motivates a weak link between BC and FC. The mathematical model aids us in finding the number of links between the two communities as a measure of stability for the large scale social structure. We study the cumulative social capital of the backward classes on discrete time steps and observe how it changes when this system is in place.

The proposed model is a modified but simple and natural model, where it is common knowledge that the state of the world changes deterministically over time, as new network connections are added through time steps. As our main contribution, we introduce in this paper the prominent role played by the strength of weak ties in alleviating the divide between the caste groups in the Indian scenario. We find it sufficient to insert a minimal number of links between the two clusters, in order to foster harmonic relations between the two conflicting groups. As a long term aim of the reservation system, we see benefits reasonably distributed evenly among members from the forward as well as the backward communities. However, the current statistics show a clear tip in the balance favoring the socially forward community, with a majority of the country's shared resources such as education, wealth, and land-holdings, being in the possession of or being accessible to only one section of the society. This undesirable disparity can be seen as the result of many recent studies, including the works of Kumar and Rustagi [9], Sedwal and Kamat [10] and Biradar [11].

The effects of the absence of the system can not be studied in the present day scenario due to the obvious reasons, however, a similar study has been performed by Borooah et al. [12]. The study took into account a social group within the country which was of the same social, educational and economic status as the Scheduled Castes (SC) and Scheduled Tribes (ST) in the pre-independence era. However, the condition of this group was observed to be at much more elevated state, proving the efficiency of the Reservation System. Many such similar comparative studies along with extant evidence only bolster the claim that this system is indeed a beacon of hope for the social disparity in India [13,14,15,16].

We also conducted a survey⁴ among people who belong to various educational institutions following reservation system. These people are affected directly by the caste based reservation system. After garnering 1005 responses from the survey, we came up with few observations that are discussed below.

⁴ The opinion survey details and results are placed in appendix D. In survey, 66.47% people are from non-reserved category and the remaining are from the reserved category.

- When we inquired the FC students regarding their meritorious opinion on BC students, they told 25.11% of BC students they interacted with had broken the stereotype associated with the reserved category students.
- The second observation was that the students from BC background, who were able to gain admission into the educational institution through reservation, are able to influence 57.57% of their younger siblings and friends from their community to try and gain admission into the institutions, causing genuine competition and increase of count on merit performance from BC background.

Both of these observations show the impact of *Forward and Backward breeze effects* caused by the reservation system. This is discussed in detail in section 3, and the mathematical model is build based on these observations.

The rest of this paper is organized as follows. Next, we discuss the required preliminaries followed by the network-based analysis of caste reservation system. The simulation results are discussed in Section 4. The paper is concluded in Section 5. The proposed model has various future directions that are also discussed in the conclusion.

2 Preliminaries and Definitions

Let $G(V, E)$ represents the undirected social network under consideration, and $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$ represent the induced subgraphs of G , where V_1 is the set of all BC nodes and V_2 is the set of all FC nodes. Therefore, $V_1 \cup V_2 = V$ and $V_1 \cap V_2 = \emptyset$. Let n_1 and n_2 be the shorthand notation for the number of individuals in the BC and FC respectively i.e. $|V_1| = n_1$ and $|V_2| = n_2$.

We define the edge set B as $(E - (E_1 \cup E_2))$ i.e. B consists of precisely those edges $\{u, v\}$, where $u \in V_1$ and $v \in V_2$, we will henceforth address these edges as *bridges*. The distance $d(u, v)$ between two nodes u and v represents the length of the shortest path between u and v . For each $u \in V_1$ i.e. a person belonging to BC, we define d_u^* as,

$$d_u^* = \min \{k | \exists v \in V_2 \ni d(u, v) = k\}$$

Therefore, d_u^* is the minimum distance from node u at which it will find at least one node of V_2 . We will refer to this parameter as the *social distance* of node u from the FC.

A path $\langle v_1, v_2, v_3, \dots, v_k \rangle$ is called an *entry path* if $v_k \in V_2$ and $v_i \in V_1 \forall 1 \leq i \leq k - 1$. Therefore, if $d_u^* = l$, then l is the length of the shortest entry path starting from node u .

3 Caste Reservation System: A Network Analysis Approach

The homophily observed in the social structure under consideration is *selection* based [2] i.e. the common characteristics that bound people together are

immutable, in this case, it is the caste of an individual. What the Reservation System does in essence is, it picks a BC individual and gets it in contact with a group of closely knit FC individuals. For example, a BC student getting a seat in a university through the reservation, implicitly creates friendship ties with a group of close FC students. The addition of such bridges has two-fold benefits, we term these the *forward breeze* effect and the *backward breeze* effect. The *forward breeze* represents the change in the mindset of the FC, on coming in contact with the BC and the *backward breeze* represents the increased motivation felt by the BC to achieve upliftment, by being influenced by FC members close to them.

Next, our aim is to calculate the gain in the social capital of the BC as a function of the bridges added in the network. There exists no universal definition or technique for measuring social capital [17], this can be attributed to the inherent subjectivity in the concept of social capital. However, in an exhaustive survey [18], the author differentiates between the social capital of different types:

1. social capital of an individual with respect to her position in the social network.
2. social capital of a group with respect to the underlying relationships within the group.
3. social capital of a group with respect to the network topological connections to other groups.

In the Caste Reservation scenario, the cumulative social capital to be calculated falls under category 3. The social capital of category 3 was first studied in [19], where the author suggests that the teams with strong outside connections generally performed better compared to groups with weaker connections outside their group. Everett and Borgatti proposed a network measure termed *group centrality* to quantize social capital of type 3 [20]. We adopt a modified version of this definition, which fits well with our model. We define the cumulative social capital of BC as the linear sum of social capital of all the individuals present in BC. Further, for every individual u in BC, we assume its social distance (d_u^*) to be a direct measure of its social capital. Lower the social distance of an individual u , higher is its social capital and vice versa.

As stated earlier, a person from BC getting reservation implies that she has an opportunity to form ties with a set of closely knit FC individuals. But for the sake of our analysis, we assume that only one tie exists per reservation i.e. all this bunch of weak ties is equivalent to one bridge while calculating the social capital of BC. This is a safe assumption, since, we are measuring social capital as a function of distance, which will rarely change for a pair of nodes in the network when we remove multiple copies of similar functioning edges. However, these multiple edges with respect to one reservation are not equivalent to one bridge in every aspect. For example, the presence of multiple weak ties amplifies the strength of the bridge across, in a sense that, even if one link breaks in the future, it does not influence the network topology or social capital significantly.

Our aim is to analyze the fall in d_u^* ($u \in V_1$) as a function of the number of random bridges added in the system. Generally, the social networks depict

scale free degree distribution [21] i.e. $P(\text{degree}(u) = k) \propto k^{-\gamma}$, where $2 < \gamma < 3$. Hence, ideally we must consider both communities to be scale free graphs. However, for making the analysis of social distance more tractable we will assume G_1 and G_2 to be *Erdos-Renyi* random graphs [22] with parameters (n_1, p_1) and (n_2, p_2) . Empirically, the social distance i.e. d_u^* falls at nearly the same rate, independent of whether we consider communities to be scale free graphs or random graphs for the same number of nodes and edges, as shown in figure 1.

Further, for every $u \in V_1$ and $v \in V_2$, the edge $\{u, v\}$ is present with probability b , which we term as the *bridging probability*. We define this new probability space of graphs as *Coupled Erdos-Renyi Graphs* where both communities are represented using two *Erdos-Renyi* random graphs and reservation links are represented as bridge edges placed between them.

3.1 Social Distance Analysis on Coupled Erdos-Renyi Graphs

The two classes BC and FC are represented by two *Erdos-Renyi* random graphs $G_1(n_1, p_1)$ and $G_2(n_2, p_2)$ respectively, where $V(G_1) = \{1, 2, 3, \dots, n_1\}$ and $V(G_2) = \{n_1 + 1, n_1 + 2, \dots, n_1 + n_2\}$. All the results proved in this paper will be for asymptotically large graphs G_1 and G_2 i.e. $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$. Every possible edge across the two graphs ($n_1 n_2$ in total) is added with the bridging probability b .

Let u represents an arbitrary node of BC. Our analysis is aimed to calculate d_u^* i.e. the social distance of node u from the FC. We begin by developing a few preliminary results.

Lemma 1. $\frac{n!}{(n-l)!} \sim^5 n^l$ as $l^2/n \sim 0$.

The proof of this lemma is provided in Appendix A.

Let M_l represents the total number of possible entry paths of length l with u as one of its endpoint. The next lemma provides an approximation for the constant M_l as a function of l .

Lemma 2. $M_l \sim n_2(n_1)^{l-1}$ as $l^2/n_1 \sim 0$.

Proof. To construct an entry path of length l with u as one of its endpoint, we need a vertex from V_2 and a sequence of $l-1$ vertices from $V_1 - u$ i.e. 1 node is to be selected from n_2 nodes and $l-1$ nodes are to be selected from $n_1 - 1$ nodes, and these $l-1$ selected nodes can be permuted in $(l-1)!$ ways.

$$\begin{aligned} \Rightarrow M_l &= \binom{n_2}{1} \binom{n_1 - 1}{l - 1} (l - 1)! \\ &= n_2 \frac{(n_1 - 1)!}{(n_1 - l)!} \\ &\sim n_2(n_1)^{l-1} \quad (\text{from Lemma 1}) \end{aligned}$$

⁵ $f(n) \sim g(n)$ if $\lim_{n \rightarrow \infty} f(n)/g(n) = 1$

□

Further, let X_l represents a random variable which is equal to the number of entry paths of length l with u as one of its endpoint. Our next result calculates the average number of entry paths of length l originating from the BC node u .

Lemma 3. $E[X_l] \sim (n_2 b)(n_1 p_1)^{l-1}$ as $l^2/n_1 \sim 0$

Proof.

$$\begin{aligned}
 X_l &= \sum_{i=1}^{M_l} Y_i \\
 \text{where } Y_i &= \begin{cases} 1 & \text{if } i^{\text{th}} \text{ entry path is present} \\ 0 & \text{otherwise} \end{cases} \\
 \Rightarrow E[X_l] &= \sum_{i=1}^{M_l} E[Y_i] \quad (\text{using linearity of expectation})
 \end{aligned}$$

An entry path $P = \langle v_{\alpha_1}, v_{\alpha_2}, v_{\alpha_3}, \dots, v_{\alpha_{l+1}} \rangle$ of length l exists if $(l-1)$ edges $(\{v_{\alpha_1}, v_{\alpha_2}\}, \{v_{\alpha_2}, v_{\alpha_3}\}, \dots, \{v_{\alpha_{l-1}}, v_{\alpha_l}\})$ are present in G_1 and the bridge $\{v_{\alpha_l}, v_{\alpha_{l+1}}\}$ is also present. Therefore, the probability that the entry path P exists, is equal to $(p_1)^{l-1}b$.

$$\begin{aligned}
 \Rightarrow E[X_l] &= M_l (p_1)^{l-1} b \\
 &\sim (n_2 b)(n_1 p_1)^{l-1} \quad (\text{from Lemma 2})
 \end{aligned}$$

□

We are interested only in the case where $b < 1/n_2$, since for $b \geq 1/n_2$, expected number of bridges per node in BC will be greater than or equal to 1, which is unrealistic in the Caste Reservation scenario. Henceforth, throughout the analysis, b is assumed to be less than $1/n_2$.

Theorem 1. *For a random graph $G_{n,p}$, $p_0 = \log(n)/n$ is the threshold probability for the property of connectedness.*

A detailed proof is available at [23].

Since the two considered graphs G_1 and G_2 are connected, the above lemma provides a lower bound on p_1 and p_2 and hence on the density of the graphs G_1 and G_2 . Therefore, $n_1 p_1 > \log(n_1)$ and $n_2 p_2 > \log(n_2)$.

Next, we analyze the quantity X_l i.e. the number of entry paths from node u as a function of l . For small values of l the number of entry paths X_l will be negligible ($\ll 1$). Our aim is to find the smallest distance d such that there exists at least one entry path of length d from node u . We prove that distance d is equal to $\log_{(n_1 p_1)}(1/n_2 b) + 1$. Henceforth, for the sake of simplicity, we will represent the quantity $\log_{(n_1 p_1)}(1/n_2 b)$ by d_0 .

Theorem 2. *The probability to exist an entry path of length less than or equal to d_0 with u as its endpoint is almost equal to zero i.e. $P(X_i = 0) \sim 0$ for $1 \leq i \leq d_0$.*

Proof.

$$\begin{aligned}
P(X_i \geq a) &\leq E[X_i]/a && \text{(using Markov's inequality)} \\
\Rightarrow P(X_i \geq 1) &\leq E[X_i] \\
&\sim (n_2 b)(n_1 p_1)^{i-1} && \text{(from Lemma 3)} \\
&\leq (n_2 b)(n_1 p_1)^{d_0-1} && \text{(Since } i < d_0) \\
&= \frac{1}{n_1 p_1} \\
&< \frac{1}{\log(n_1)} && \text{(from Theorem 1)} \\
\Rightarrow P(X_i \geq 1) &\rightarrow 0 \\
\Rightarrow P(X_i = 0) &\rightarrow 1
\end{aligned}$$

□

Further, we will prove that almost always (i.e. with probability close to one) there exists at least one entry path of length $d_0 + 1$ from u . Hence, proving our claim that $d_u^* = d_0 + 1$.

Lemma 4. *For any random variable X , $P(X = 0) \leq \frac{\sigma_X^2}{\mu_X^2}$, where σ_X and μ_X represent the variance and mean of the random variable X respectively.*

Its proof is discussed in Appendix B.

Lemma 5. *The standard deviation of the random variable X_{d_0+1} approaches zero i.e. $\sigma_{X_{(d_0+1)}} \sim 0$*

Proof.

$$\begin{aligned}
X_l &= \sum_{i=1}^{M_l} Y_i \\
\Rightarrow X_l^2 &= \sum_{i=1}^{M_l} \sum_{j=1}^{M_l} Y_i Y_j = \sum_{k=0}^l Z_k
\end{aligned}$$

where Z_k accounts for all $Y_i Y_j$'s, where the i^{th} and j^{th} entry paths have precisely k edges in common. Let $|Z_k|$ represent the number of terms in Z_k 's summation.

$$|Z_0| \geq \binom{n_1-1}{l-1} (l-1)! \binom{n_2}{1} \binom{n_1-l}{l-1} (l-1)! \binom{n_2-1}{1}$$

If none of the vertices in the two entry paths are common, then certainly none of its edges are common either, this gives us the above inequality.

$$\begin{aligned} \binom{n_1-1}{l-1} (l-1)! \binom{n_2}{1} \binom{n_1-l}{l-1} (l-1)! \binom{n_2-1}{1} &= \frac{(n_1-1)!}{(n_1-2l+1)!} n_2 (n_2-1) \\ &\sim \frac{n_2(n_2-1)}{n_1(n_1-2l+1)} n_1^{2l} \\ &\sim n_2^2 n_1^{2l-2} \end{aligned}$$

The total number of terms in the summation of X_l^2 are M_l^2 i.e. approximately $n_2^2 n_1^{2l-2}$. Therefore, most of the summation terms of X_l^2 fall into the basket of Z_0 .

$$\begin{aligned} E^2[X_{d_0+1}] &\sim 1 && \text{(from Lemma 3)} \\ \sigma_{X_l}^2 &= E[X_l^2] - E^2[X_l] && \text{(by definition)} \\ \implies \sigma_{X_{(d_0+1)}}^2 &= E[X_{d_0+1}^2] - E^2[X_{d_0+1}] \\ \implies \sigma_{X_{(d_0+1)}}^2 &\sim 0 \end{aligned}$$

□

Theorem 3. *Almost always there exists an entry path of length equal to $d_0 + 1$ with u as its endpoint i.e. $X_{d_0+1} \geq 1$.*

Proof.

$$\begin{aligned} P(X_{(d_0+1)} = 0) &\leq \frac{\sigma_{X_{(d_0+1)}}^2}{\mu_{X_{(d_0+1)}}^2} && \text{(using Lemma 4)} \\ \implies P(X_{(d_0+1)} = 0) &\sim 0 && \text{(using Lemma 3 and 5)} \\ \implies P(X_{(d_0+1)} \geq 1) &\sim 1 \end{aligned}$$

□

Therefore, almost always $d_u^* = \log_{n_1 p_1}(1/(n_2 b)) + 1$. Since, this formula is independent of u , almost all nodes in the BC have social distance $(d_0 + 1)$. Let x represents the expected number of bridges added in the system.

$$\begin{aligned} \implies x &= n_1 n_2 b \\ \implies d_i^* &= \log_{n_1 p_1}(n_1/x) + 1 \end{aligned}$$

$$d_i^* = \frac{\log(n_1) - \log(x)}{\log(n_1 p_1)} + 1 \quad (1)$$

The above theorem proves that the social distance of any arbitrary node i reduces logarithmically as a function of the number of bridges in the system. Therefore, only the first few bridges are highly effective in reducing the distance between the two communities, and the bridges that are added later on, don't bring a significant change in the social capital of an individual present in BC.

4 Simulation Results

4.1 Network Models

We have used following synthetic network generative models to study the impact of adding bridges on the social distance between both communities.

1. Erdos-Renyi (ER) Model: In 1969, Erdos and Renyi proposed a model to generate random networks [22]. In this model, there are n nodes and an edge is placed between a pair of nodes with some fixed probability p . The mathematical analysis of the proposed model is explained for such type of networks.
2. Barabasi-Albert (BA) Model: In 1999, Barabasi and Albert observed that real world networks are not random but scale-free⁶ [21]. They are based on the rich-gets-richer phenomenon, where a node having higher degree has a high probability of getting new connections. The degree distribution of scale-free networks follows power law, so the probability of a node having degree k is defined as $ck^{-\gamma}$, where c is a constant and γ is the power law exponent.

To simulate the proposed model, both FC, as well as BC community, are generated using the same model and having the same properties like network size, density, etc.

4.2 Discussion

In this section, firstly, we discuss the verification of math model using simulation. Secondly, we study how the effective social distance is reduced for different types of networks.

The verification of math model is shown in Figure 1, where x-axis shows the number of bridges and y-axis shows the average social distance. To compute the average social distance, first the social distance is computed for each node of BC and then its average is taken. The experiment is repeated 10 times to compute the average social distance for the different number of bridges. In figure 1, red color shows the average social distance computed using equation 1 of the proposed model and blue color shows the average social distance computed using simulation on coupled Erdos-Renyi networks.

In the math model, the average social distance is converged to 1 when n_1 bridges are placed as each BC node will be directly connected to one FC node. But in coupled Erdos-Renyi network, the average social distance is not converged to 1 (close to 1) when n_1 bridges are placed as the bridges are placed uniformly at random and one node of BC can be connected with multiple nodes of FC.

As we know, real-world networks possess scale-free structure, next, we study the average social distance on scale-free BA networks. The results are shown in figure 1, where the average degree of the network is same as the corresponding ER networks. The results show that the scale-free networks follow the same pattern

⁶ The detailed model is explained in appendix C.

for average social distance and it decreases logarithmically. The mathematical analysis of these models is left as the future work. The similar results are obtained for the networks of different sizes and densities.

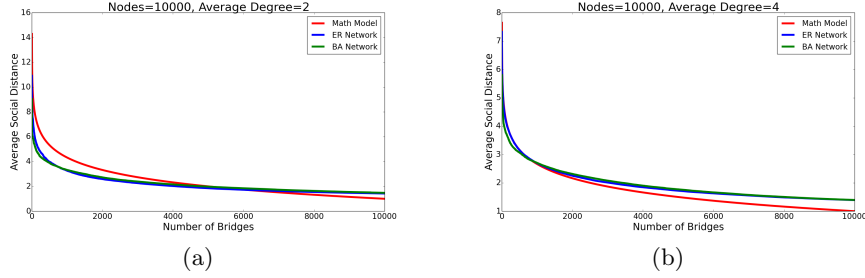


Fig. 1. Average social distance versus number of bridges

Real-world networks possess meso-scale structures like community and core-periphery structure [24,25]. We also simulate the proposed model on the following real-world networks.

1. Facebook Network: Facebook is the most popular online social networking website today. This dataset is the induced subgraph of Facebook [26], where users are represented by nodes and friendships are represented by edges. It contains 63,392 nodes and 816,831 edges.
2. Twitter Network: This is an induced subgraph of Twitter [27]. Each node is a Twitter user, and each directed edge from user A to user B means that user A follows user B. This is converted into undirected network for the study and it contains 81,306 nodes and 1,342,296 edges.

To simulate the proposed model, two copies of real-world networks are created to represent BC and FC community. The results for Facebook and Twitter networks are shown in Figure 2. The average social distance of real-world network is also compared with the math model by taking the same number of nodes and network densities. The results show that the average social distance decreases logarithmically in real-world networks. The bridges are placed randomly and a node can be connected with multiple bridges, so, the distance is not converged to 1 after placing n_1 number of bridges where n_1 is the size of BC community.

The simulation results support the proposed model. The results prove that the impact of placing more bridges decreases with time and a small number of bridges are sufficient to maintain the harmonic distance between the communities.

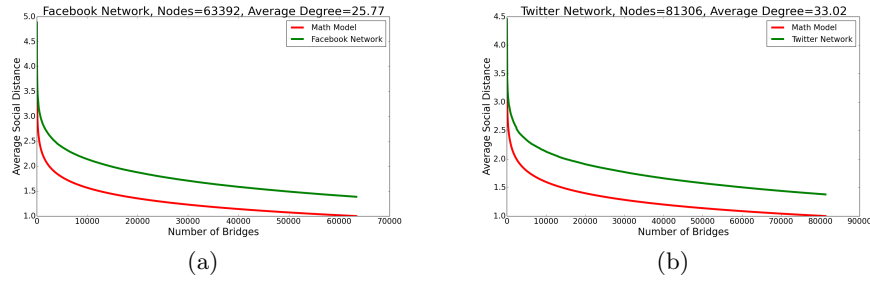


Fig. 2. Average social distance on real world networks, a. Facebook, and b. Twitter

5 Conclusion and Future Directions

The Indian society has suffered for a long time due to the discriminatory system of caste-based segregation that initially arose from the concepts of spiritual purity. Indian government established the reservation system to provide equal opportunities in education and employment to classes that have historically been denied access to the resources. In our paper, we looked at the reservation system from a purely network theoretic perspective, by modeling the polarized Indian society in the form of a homophilic network and considering reservation to be the phenomenon by which link formation between the two polar groups is initiated.

We defined the social capital associated with each individual in the backward class as a function of its social distance to forward class, that quantifies an individual's access to education and employment opportunities. In the proposed model, we studied the increase in social capital of a member of the disadvantaged group as a function of the number of inter-group links (reservation opportunities). We noted that a very small number of links between the two groups are enough for the cumulative benefit to increase rapidly. To the best of our knowledge, such a model of the reservation system is the first of its kind.

As a part of future work, we plan to investigate a larger variety of social applications wherein such a disparity exists, and apply a similar system in order to test its efficiency. This will allow us to further analyze different parameters involved in the cumulative social capital of a group. Additionally, we plan on studying the social structure within the Indian subcontinent in greater detail, and arrive at a specific percentage of reservation, called as the *Ideal/Optimum Number*, which if applied, will cause great amount of uplift within shortest time, and will balance the distribution of resources between both groups.

We will further analyze the changes in the opinion of people towards other community that is necessary for the harmonic existence of a society. This also has been the one important motive of the reservation system. We would like to propose a mathematical model to study this phenomenon and the opinion formation in the proposed model. This will help to understand the harmonic stability in the society due to the opportunities given by the Reservation System.

References

1. Susan Bayly. *Caste, society and politics in India from the eighteenth century to the modern age*, volume 3. Cambridge University Press, 2001.
2. Miller McPherson, Lynn Smith-Lovin, and James M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:pp. 415–444, 2001.
3. Mark S Granovetter. The strength of weak ties. *American journal of sociology*, 78(6):1360–1380, 1973.
4. David Strang and Sarah A Soule. Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual review of sociology*, 24(1):265–290, 1998.
5. Ivo Welch. Sequential sales, learning, and cascades. *The Journal of finance*, 47(2):695–732, 1992.
6. Abhijit V Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817, 1992.
7. Akshati Saxena, SRS Iyengar, and Yayati Gupta. Understanding spreading patterns on social networks based on network topology. In *Advances in Social Networks Analysis and Mining (ASONAM), 2015 IEEE/ACM International Conference on*, pages 1616–1617. IEEE, 2015.
8. Matthew O JACKSON. Social structure, segregation, and economic behavior. presented as the nancy schwartz memorial lecture, 2007.
9. AK Shiva Kumar, Preet Rustagi, et al. Elementary education in india: Progress, setbacks, and challenges. Technical report, 2016.
10. Mona Sedwal and Sangeeta Kamat. Education and social equity with a special focus on scheduled castes and scheduled tribes in elementary education. 2008.
11. RR Biradar. Social and economic change monographs. 2011.
12. Vani K Borooah, Amaresh Dubey, and Sriya Iyer. The effectiveness of jobs reservation: caste, religion and economic status in india. *Development and change*, 38(3):423–445, 2007.
13. Thomas E Weisskopf. Impact of reservation on admissions to higher education in india. *Economic and Political Weekly*, pages 4339–4349, 2004.
14. Karuna Chanana. Accessing higher education: the dilemma of schooling women, minorities, scheduled castes and scheduled tribes in contemporary india. *Higher Education*, 26(1):69–92, 1993.
15. Binod Khadria. Human resources in science and technology in india and the international mobility of highly skilled indians. 2004.
16. Viney Kirpal and Meenakshi Gupta. *Equality through reservations*. Rawat Publications, 1999.
17. Philippe Aghion and Steven Durlauf. Preface to the handbook of economic growth, 2005.
18. Stephen P Borgatti, Candace Jones, and Martin G Everett. Network measures of social capital. *Connections*, 21(2):27–36, 1998.
19. Deborah Gladstein Ancona. Outward bound: strategic for team survival in an organization. *Academy of Management journal*, 33(2):334–365, 1990.
20. Martin G Everett and Stephen P Borgatti. The centrality of groups and classes. *The Journal of mathematical sociology*, 23(3):181–201, 1999.
21. Albert-Laszlo Barabasi and Reka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
22. Paul Erdős and Alfréd Rényi. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

23. Béla Bollobás. *Modern graph theory*, volume 184. Springer Science & Business Media, 2013.
24. Akarti Saxena and SRS Iyengar. Evolving models for meso-scale structures. In *Communication Systems and Networks (COMSNETS), 2016 8th International Conference on*, pages 1–8. IEEE, 2016.
25. Yayati Gupta, Akarti Saxena, Debarati Das, and SRS Iyengar. Modeling memetics using edge diversity. In *Complex Networks VII*, pages 187–198. Springer, 2016.
26. Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P. Gummadi. On the evolution of user interaction in facebook. In *Proceedings of the 2Nd ACM Workshop on Online Social Networks, WOSN '09*, pages 37–42, New York, NY, USA, 2009. ACM.
27. Jure Leskovec and Julian J Mcauley. Learning to discover social circles in ego networks. In *Advances in neural information processing systems*, pages 539–547, 2012.

Appendix

Appendix A Proof of Lemma 1

Proof.

$$\begin{aligned}
\frac{n!}{(n-l)!} &\sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \frac{1}{\sqrt{2\pi(n-l)}} \left(\frac{e}{(n-l)}\right)^{n-l} \quad (\text{using Stirling's approx.}) \\
&\sim \left(\frac{n}{e}\right)^l \left(1 - \frac{l}{n}\right)^{-(n-l+\frac{1}{2})} \\
&\sim \left(\frac{n}{e}\right)^l (e)^{l(n-l+\frac{1}{2})/n} \\
&\sim n^l
\end{aligned}$$

□

Appendix B Proof of Lemma 4

Proof.

$$\begin{aligned}
P(|X - \mu_X| \geq a) &\leq \frac{\sigma_X^2}{a^2} \quad (\text{Chebyshev's inequality}) \\
\implies P(|X - \mu_X| \geq \mu_X) &\leq \frac{\sigma_X^2}{\mu_X^2} \\
P(X = 0) &\leq P(|X - \mu_X| \geq \mu_X) \\
&\leq \frac{\sigma_X^2}{\mu_X^2}
\end{aligned}$$

□

Appendix C Barabasi-Albert Model

In 1999, Barabasi and Albert observed that real-world networks are not random. They observed that real-world networks follow power law degree distribution that indicates that there exists very few nodes having higher degrees and most of the nodes having lower degrees. The real world networks having power law degree distribution are also called scale-free networks. In scale-free networks, the probability $f(k)$ of a node having degree k is defined as,

$$f(k) = ck^{-\gamma} \quad (2)$$

where, γ is the power law exponent, and for real-world scale-free networks its range is $2 < \gamma < 3$.

Based on this observation, they proposed an evolutionary preferential attachment model to generate synthetic networks that follow the properties of real-world scale-free complex networks [21]. This model starts with a seed graph having n_0 nodes that are connected with each other. At each time stamp, a new node is added to the network and it makes connections with m already existing nodes. The probability $\Pi(u)$ of an existing node u to get a new connection is directly proportional to its degree $deg(u)$. It is defined as,

$$\Pi(u) = \frac{deg(u)}{\sum_v deg(v)}$$

So, the nodes having higher degrees acquire more links over time, and the degree distribution is skewed towards lower degrees. As the network grows, only a few nodes called hubs manage to get a large number of links.

Appendix D Survey Details

We conducted a survey⁷ amongst the very people directly involved and are affected by caste based reservation system. We gathered opinion from both ends, students who have availed the reservation and general merit students. After garnering 1005 responses from the survey conducted across various educational institutions following reservation system, we came up with few observations corresponding to our result. In our survey, we observe how much change has already swept over both communities from either side, which is the direct result of the reservation system. All people who participated in survey belong to age 17-36. Below are the questions that we asked in the survey:

- Please select type of institution where you have studied.
 1. IIT/NIT/IISERS/govt Institution
 2. Semi Government

⁷ The survey form is available at

<https://docs.google.com/a/iitrpr.ac.in/forms/d/1afOga9PbEQeNe3wYVFHKQ1f8JcAoyYIuOlXW3lAJD2A>

3. Private

- Have you ever availed the reservation? Choose one of the options.
 1. No
 2. Yes (caste-reservation/management quota)
- Average number of relatives (family members) do you see in an year.
- The percentage of family members that you can influence strongly with your opinion. (answer in %)?
- Question for NON-RESERVED Category Students: What percentage of reserved category (caste-reservation/management quota) students have impressed you with their skills?(answer in %)?
- Question for RESERVED Category Students: The overall percentage of your relatives (younger ones) who look up to you for inspiration to try and achieve like you. (answer in %)?
- Question for ALL Students (Answer it based on your view): What percentage of your family members mentioned above, can influence their friends and connections with your opinion (your academic story)? (answer in %)?