ENVISAGING PROMINENCE OF INDIAN TELECOM OPERATORS USING AN ENSEMBLE LINK BASED APPROACH

Amulyashree Sridhar
Research Scholar, Department of Computer Science and Engineering,
RV College of Engineering, affiliated to Visvesvaraya Technological University,
Belgaum, Karnataka 590018, India.
0908amulyashree@gmail.com

Sharvani GS
Associate Professor, Department of Computer Science and Engineering,
RV College of Engineering, affiliated to Visvesvaraya Technological University,
Belgaum, Karnataka 590018, India.
sharvanigs@rvce.edu.in

AH Manjunatha Reddy
Associate Professor, Department of Biotechnology,
RV College of Engineering, affiliated to Visvesvaraya Technological University,
Belgaum, Karnataka 590018, India.
ahmreddy@gmail.com

Kalyan Nagaraj
Research Scholar, Department of Computer Science and Engineering,
RV College of Engineering, affiliated to Visvesvaraya Technological University,
Belgaum, Karnataka 590018, India.
kalyan1991n@gmail.com

Abstract - Understanding ambiences from consumers helps in inspecting market value of an artifact. Data and voice quality of telecom service are often praised and targeted in social networking platforms. There is a need to analyze these opinions to infer renown of service providers. In this context, data is extracted for major Indian telecom service providers from Twitter and FaceBook. Furthermore, performance metrics of these providers are collected from TRAI portal to compare projected thresholds with customer opinions. These datasets are retrieved in predefined timeframe to ensure that information isn't biased. Towards exploring a prevalent service provider, an ensemble node ranking approach is designed based on ideologies of SVM and RNN algorithms (SVMRNNRank). This approach is modelled based connectivity in the network. Its effectiveness is demonstrated by calculating certain statistical metrics. Comparative analysis reflects that influential nodes identified from SVMRNNRank have better acceptance amongst social network users and TRAI performance indices.

***Keywords*: Telecom service providers; node ranking; SVMRNNRank.

1. Introduction

Evolution of digital era has paved way for growth of immense information over the last decade. Several sectors have predominantly contributed to this global data explosion. According to a survey conducted by IDC, it is reported that worldwide data would proliferate upto 175 zettabyte by 2025 [Reinsel (2018)]. Concurrently, digitalization in Indian market has contributed to increased productivity among multiple sectors. Telecommunication is one arena that has attributed to remarkable progress owing to paradigm shift from wired to wireless transmission [Gupta (2018)]. This advancement has in turn led to development of advanced smart phone devices at user-friendly prices. Proliferation of these devices has largely contributed to progression of internet from second-generation communication (2G) to advanced fifth generation (5G) communication [Liu (2016)]. These expansions have paved way for unrestricted internet access across all arenas in the country. According to the report released by Internet and Mobile Association of India (IAMAI), the nation estimates about 451 million active subscribers at the end of 2019 [IAMAI (2019)]. Such massive data growth in telecommunication sector is projected to contribute about 8.2 % to the Gross Domestic Product (GDP) by 2020 [Key Highlights of Economic Survey 2018-19].

This digital data explosion has thereby resulted in immense competition among telecom service providers to deliver unwavering data and voice coverage reciprocally in urban and rural arenas. Service providers are in constant urge to satisfy customer's requirements at numerous echelons. They proclaim exciting offers at periodic intervals to avoid service breach by customers [Rajeshwari (2017)]. In coming years, it would turn out to be even more perplexing for these operators to promote advanced infrastructure for facilitating the ever increasing data demand from customers [Deo (2017)]. Henceforth, understanding the dynamics of user preferences will help in deciphering better strategies to balance acquisition and retention of customers in the long run [Yadav (2019)].

Customer opinions are usually dynamic and heterogeneous in nature, making it further challenging to deduce valuable insights [Hidayati (2018)]. In this perspective, it is constructive to represent these sentiments as a network structure to comprehend inherent relationships among multiple entities. Discovering such associations would moreover unveil virtues and deficiencies of a network operator which might have led to approbation or renunciation of its service by the public. Furthermore, capturing these behavioral indications by orienting them as structural graphs broadly falls into the category of social networks. A social network comprises of nodes and their associations as links [Pinheiro (2011)]. In this case, telecom operators are designated as nodes while their connotations are symbolized as links. These operators and their links are demarcated as telecom network graph. Several studies in past have focused on implementation of social network analysis in telecommunication sector pertaining to different applications including call records modelling [Mishra (2018)], anomaly detection [Chaparro (2015)], product marketing [Insani (2016)] and customer satisfaction [Andresen (2017)]. However, these studies have not explored the importance of link structures in ranking a node as prevailing or recessive. Understanding these physiognomies using links is attributed to the field called 'Link Mining'. This field has multidisciplinary origins structured on web mining, relational learning and information extraction [Getoor (2005)]. Link mining techniques are designed to support node based activities, link based activities and network based activities. Numerous studies have focused on implementing link based strategies for supporting advanced knowledge extraction from unstructured and semi-structured data [Lim (2019); Güven (2019)]. Owing to its diverse applications, link mining techniques could be adopted to decipher the impact of telecom operators in current scenario.

Grounded on these principles, the current study is an attempt to apprehend impression of telecom service providers among customers. Primarily, sentiments pertaining to telecom operators are extracted from popular social networking platforms including Twitter and FaceBook. Following data extraction, network structure is generated for each operator to disclose substantial emotions. Furthermore, significant features within the network are identified by probing associations among link structures. Centered on the network derived, prominence of each operator is predicted using conventional node ranking approaches. Amongst the adopted approaches, a best performing node ranker which precisely groups the telecom operators is selected. Moreover, predictions from the approach are validated using statistical approaches to identify best performing service provider for the current timeframe.

2. Proposed approach

This section describes the procedures adopted for discovering substantial telecom operators in Indian market.

2.1. *Telecom data collection*

Among manifold variants of telecommunication data, it is often difficult to access specific information owing to stringent guidelines from regulatory bodies. In this context, it is relevant to consider a sample set for modelling operator dynamics rather than considering an entire population of diverse information. Henceforth, dynamics of telecom operators in wireless mode is considered in this study owing to its popularity compared to its wired opponent. Two samples from wireless telecommunicating are considered in this study. The data is described as follows:

2.1.1. *Social network data (S_1)*

Popular social networking platforms including Twitter and FaceBook are selected to capture real sentiments from customers. A timeframe is chosen for data collection to avoid predisposed opinions owing to upgradation, confiscation and promotion of multiple plans by telecom service providers. Data is gathered from these online podiums for four telecom operators dominant in Indian market namely Bharti Airtel, Reliance Jio, VodafoneIdea and Bharat Sanchar Nigam Limited (BSNL). Keyword search is employed to accumulate customer sentiments. For instance, customer reactions for Airtel operator are captured by using "#Airtel" and "Airtel" as keywords in Twitter and FaceBook respectively. For each operator, data in form of tweets and comments are accrued within a proposed timeframe between November 2019 and March 2020. R programming language is used to collect data from these podiums using in-built dependencies 'rtweet' [Kearney (2020)] and

'Rfacebook' [Barbera (2017)]. The data so derived from these platforms are labelled as S_{1T} for Twitter data and S_{1F} for facebook data.

2.1.2. *TRAI wireless performance data (T_1)*

Similar to previous data sample, performance statistics of telecom operators in wireless mode is collected from Telecom Regulatory Authority of India (TRAI) web portal [TRAI (2019)]. Different parameters like wireless subscribers and quality of service statistics are extracted from the repository. These parameters are extracted in the same timeframe (i.e. November 2019 to March 2020) as that of social network data to maintain uniformity. Suppose there is no pre-existing data available at this timeframe, previously updated data is selected for all the telecom operators from the portal. The data so derived from this platform is labelled as T_1. The latest report corresponding to the defined timeframe is selected for the study.

2.2. *Data preprocessing*

2.2.1. *S_1*

Data collected from social networking platforms (i.e. S_{1T}, S_{1F}) needs to be pre-processed to confiscate inappropriate context. Information mined from these sites are subjected to initial cleaning for eliminating weblinks, hashtags, numbers, english stopwords and punctuation symbols. From the derived data white spaces are removed and stemming is performed to reduce the words to their base form. All these tasks are performed in R programming language using the inbuilt dependency 'tm' [Feinerer (2019)].

2.2.2. *T_1*

Data collected from TRAI portal is converted to .csv files and stored based on the monthly timeframe. The attributes having no predefined values are equalized to null values prior to analysis. These reports require no further preprocessing.

2.3. *Generating telecom network*

Once data samples are processed, they are represented as network structure comprising of nodes and edges. However, the denotations of these networks differ for both data samples.

2.3.1. *S_1*

The network comprises of the four telecom operators as nodes (*n*) and social network users (i.e. from Twitter and FaceBook) as links (*l*). Nodes are labelled as different telecom operators i.e. Bharti Airtel (n_1), Reliance Jio (n_2), VodafoneIdea (n_3) and BSNL (n_4). The networks so derived from S_{1T} and S_{1F} (N_{1T}, N_{1F}) are undirected in nature as the links are devoid of any direction. These undirected, unweighted networks are defined as a function of nodes and links i.e. N_{1T} (*n*, *l*). Furthermore, different node based metrics including degree, clustering coefficient, eigenvector centrality, cosine and jaccard metrics are computed to detect interrelationship among the nodes [Gómez (2013)]. The definitions are enlisted below:

- Degree Centrality: It denotes the number of links that are adjacent to a particular node of interest in an undirected graph. Greater the degree of a node, higher is its importance. It is represented mathematically as follows for node '*j*':

  $$\text{Degree (j)} = \sum_k l(j,k) \tag{1}$$

  Here, *l (j, k)* will be equated to one if there is an adjacent link between the nodes *j* and *k*. Its value will be zero if there is no edge.

- Clustering coefficient (CC): It is an estimate of number of nodes that are neighbors to each other. It is derived from local clustering of every node in a network. Local clustering for a node '*j*' is defined as:

  $$CCj = \frac{\text{Number of nodes connected by neighbors}}{\text{Total number of nodes in network}} \tag{2}$$

  The coefficient ranges from zero to one. Higher the value of coefficient, greater is the connectivity among nodes.

- Eigenvector centrality: It is often referred as an extension of degree centrality. It measures a node's influence based on its connections with other nodes. Higher the value better is the connectivity of a node.

- Cosine and jaccard metrics: Cosine metric is used to detect the extent of similarity among any two nodes in the network. Its value ranges from zero to one. Larger the value, greater is the similarity. For two nodes '*j'* and '*k'*, cosine metric is defined as:

  $$\cos(j,k) = \frac{j \times k}{\|j\|\|k\|} \tag{3}$$

Jaccard metric is the ratio of mutual neighbors between any two nodes in a network to total neighbors between the two nodes. Total neighbors are represented by union operation between the two nodes. It is represented mathematically for the nodes *j* and *k* as follows:

$$\text{Jaccard(j,k)}=\frac{\text{Neighbors(j)} \cap \text{Neighbors(k)}}{\text{Neighbors(j)} \cup \text{Neighbors(k)}} \qquad (4)$$

The data samples (S_{1T}, S_{1F} and T_1) are visualized using Gephi software version 0.9.2 [Bastian (2009)].

2.3.2. T_1

Performance statistics collected from TRAI portal is visualized as a network. Nodes are represented by telecom service providers while links represent different performance parameters. Some of the noteworthy performance parameters include service activation, successful data transmission for download, successful data transmission for upload, minimum download speed, average throughput for packet data, latency, PDP context activation success rate and drop rate. This data is further used for comparative analysis post node ranking.

2.4. *Initial node ranking and link reduction*

S_1 is taken as data for node ranking. The network derived from Twitter and FaceBook comprises of four nodes with 40,449 and 64,567 links respectively. Node ranking is performed on network data using conventional node ranking algorithms, PageRank [Page (1998)] and Hyperlink Induced Topic Search (HITS) [Kleinberg (1999)]. However, these algorithms could not differentiate nodes based on their importance. Both node rankers resulted in ranking all four nodes as prominent for both S_{1T} and S_{1B} datasets. This might be due to the complexity of the networks with fewer nodes and numerous links. Henceforth, the networks must be simplified by eliminating inappropriate links. In this direction, multiple links arising from the same source node reaching to the same destination node are eliminated. Furthermore, links having the same source as well as destination node which in turn form loops are also eliminated. Once loops and multiple edges are removed, the network size is to be further reduced to retain momentous links. In this context, link reduction techniques are to be employed such that network connectivity is unaltered. Of numerous feature selection techniques available, data mining based algorithms are preferred owing to their better performance [Kim (2003)]. In this context, a wrapper based random forest classifier boruta is applied on social network and TRAI datasets to identify prominent connections [Kursa (2018)]. The algorithm available in R programming language calculates importance score of each link attribute and eliminates weak associations iteratively. It converges once all the significant link features are identified. This step ensures that the resulting network is devoid of irrelevant connotations.

2.5. *Node ranking*

The networks derived after link reduction from S_1 are subjected to node ranking procedure to identify relevant objects. The approach adopted for node ranking is formulated based on these principles:

a) Given S_{1T}, S_{1F} in the form of $N(n, l)$, such that $\{n , l\} \varepsilon N$ where $n= \{n_1, n_2, n_3, n_4\}$; $l= \{l_1,l_2, l_3, l_4\ldots l_n\}$; Evaluate the significance of every node in the networks by randomly assigning each node as source (*s*) and destination (*d*) node.

b) Calculate the significance score of the source node such that its value is non-negative. Higher the significance score, greater is the importance of a node.

c) Iterate the steps a) and b) by randomly assigning the source node until the significance score reaches convergence

In this perspective, PageRank and HITS algorithms are re-evaluated on the network datasets to identify the prominent telecom players. The results of these algorithms differ significantly after network size is reduced. In case of both S_{1T} and S_{1F} ranking of nodes are drastically increased compared to previous scores. However, these approaches are not appropriate in ranking nodes precisely as they have a tendency to overlook structural transformations in the networks. Hence, it is imperative to rank these nodes using other approaches.

Data mining methods are suited as better alternatives in this perspective [Mariani (2015)]. Some popular algorithms adopted for node ranking in several studies include Support Vector Machine (SVM), artificial neural networks, deep learning, naïve bayes and many more. Scrutinizing a best ranking approach is essential to identify key nodes from networks. In this direction, SVM is initially chosen to rank the nodes from the networks. The variant of SVM adopted for node ranking is called Ranking SVM. This approach generalizes SVM to perform node ranking [Herbrich (2000)]. The classifier ranks the nodes by transforming them into pairs of objects. For instance, data in two social networks i.e. S_{1T} and S_{1F} have objects n_1, n_2, n_3 and n_4 in two different levels based on the link orientation. The weight function *w* is defined in linear fashion as a function of the nodes to be ranked i.e. $f(n) = \langle w, n\rangle$. This function transmutes nodes into vector space and ranks them based on their projections. Furthermore, objects within same data groups are defined as vector features by the algorithm. Labels are assigned to these features in vector space to aid in ranking of prominent and non-prominent nodes. For instance, (n_1-n_2) is considered as positive value for ranking while (n_2-n_1) is considered negative. The SVM based ranker defines hyperplane which surpasses source, significant and irrelevant nodes from network data. Followed by hyperplane, margin is represented as the minimal distance between transformed node pairs in sample and vector space. SVM in linear mode is trained in this feature space to rank network nodes based on their hyperplane, margin and weight functions. Ranking is achieved by considering weight function in Quadratic Programming (QP) mode as follows:

$$\begin{aligned}
&\min_{w,\varepsilon} \frac{1}{2}\|w\|^2 + C \sum_{k=1}^{4} \varepsilon_k \\
&\text{such that } \mathrm{y}_k \left\langle w, n_k^{(1)} - n_k^{(2)} \right\rangle \geq 1 - \varepsilon_k, \\
&\varepsilon_k \geq 0, k = 1, 2, 3, 4
\end{aligned} \tag{5}$$

Here, $n_k^{(1)}$ and $n_k^{(2)}$ represents the first and second nodes in vector space, k denotes the estimates of training data instances, C being the regularization coefficient having non-zero value, while $\left|\cdot\right|$ represents the *L2* normalization index. The deduced objective function is equivalent to minimized hinge loss regularization function. This algorithm when employed on S_{1T} and S_{1F} ranks the nodes orderly. However, there is some amount of marginal difference amongst certain nodes in networks. Despite its improved ranking compared to previous approaches, there is a need to further optimize ordering of nodes.

In this direction, deep learning architectures are suited to learn from trained models [Dahl (2012)]. Specifically, recurrent neural networks (RNNs) are appropriate as they perform predictions based on current data and previous outcomes. This sequential architecture helps in improvised node ranking by learning from trained SVM which is in turn fed to RNN. Furthermore, the ensemble SVM based RNN (SVMRNNRank) will help in improving effectiveness of node ranking. Implementation of this ensemble approach for node ranking is as follows:

Step 1: The input vector comprising of ranked nodes along with their association $\{n_k, l\}$ is considered. It is ensured that all the vectors are in the same dimension i.e. $n_k \in \mathbb{R}^{lxdm}$. RNN usually processes input in binary tree format comprising of root node (n_r) and child nodes (n_{c1}, n_{c2}) i.e. ($n_r \rightarrow n_{c1}$, n_{c2}). The root and child nodes are designated randomly and get updated iteratively. These nodes are represented in bottom-up fashion using the equation:

$$z(n_r) = a_f\left(\left[z(c_1), z(c_2)\right] \times T_m\right) \tag{6}$$

Here, a_f denotes the activation function while T_m is the transition matrix such that $T_m \in \mathbb{R}^{2d_m x 2d_m}$. Furthermore, vectors representing a node are denoted by $z\ (.) \in \mathbb{R}^{l\ x\ dm}$. The input vector $z\ (n_r,\ l)$ is further transformed into hidden features $z\ (n_h)$ by adding the projection layer such that $n_h \in \mathbb{R}^{l\ x\ dm}$. This layer includes the node ranking estimates from trained SVM. It is represented as:

$$n_h = HTh(n_r \mathrm{X} P_m) \tag{7}$$

Here, *HTh* indicates the HardTanh estimate as the activation function having its range between -1 and 1, P_m is the projection matrix such that $P_m \in \mathbb{R}^{kr\ x\ dm}$. Deriving these formulations generates a tree structure from input vector.

Step 2: Node ranking is performed by calculating the salience score S for non-terminal vertices (n) from the set of all nodes $N = \{n\}$ based on principles of regression. The strategy adopted differs for nodes based on their orientation. If a node is pre-terminal (N_p) then ROGUE-1 (RG_1) score is used for evaluation of node importance. Suppose a node is towards the root then both ROGUE-1 (RG_1) and ROGUE-2 (RG_2) scores are adopted. Based on this assumption, the salience score is defined as follows:

$$s(n) = \left\{ \begin{matrix} RG_1,\ n \in N_p \\ \beta RG_1\ +\ (1{-}\beta)\ RG_2(n),\ n \in N{-}N_p \end{matrix} \right\} \tag{8}$$

The coefficient β is set to its maximal value of 0.5 to identify importance of nodes. Suppose a parent node is affirmed as salient, its children are also said to be salient. However, the reverse isn't true and needs to be confirmed.

Step 3: Regression is performed and error function is computed. The step is reiterated until error function reaches convergence.

By following these steps, node ranking are derived SVM based RNN learner. This ensemble approach (SVMRNNRank) is able to rank the nodes and its performance is to be assessed.

2.6. *Validating node ranking approaches*

The node rankings generated by PageRank, HITS, SVM and ensemble approach are to be evaluated using statistical metrics to identify their effectiveness in the long run. As centrality metrics fails to discriminate the nodes, other metrics are to be demarcated. These metrics are defined as follows:

2.6.1. Precision (P_j): It is formulated as relevance of a node with respect to all the nodes present in the network. Its value ranges between zero and one. Higher the value better is the importance of a node.

$$P_j = \frac{\text{Relevance of a node in network}}{\text{Total number of nodes in network}} \tag{9}$$

2.6.2. Mean average precision (*MAP*): It is the average precision value estimated for all the nodes in the network. Its value also ranges between zero and one. Greater value of *MAP* indicates higher prominence of a node in the network.

$$MAP_j = \frac{\sum_{j=1}^{n}\left(P_j \times rel(n)\right)}{\text{Total number of relevant nodes in network}} \tag{10}$$

2.7. *Comparing S_1 and T_1 metrics*

Once node ranking is performed and relevant telecom players are identified from S_{1T} and S_{1F}, T_1 metrics are to be analyzed for estimating inherent correlation between social network opinions and threshold estimates. If the correlation is better, then the nodes ranked are relevant with calculated estimates. After comparison, significant telecom players are identified from S_1 and T_1.

3. Results and Discussion

This section describes the outcomes derived from telecom network analysis in identifying prominent telecom players.

3.1. *Data generation and network analysis*

Telecom datasets collected from social network platforms (S_{1T}, S_{1F}) and TRAI web portal (T_1) for wireless mode are stored as .csv files. Once data is gathered they are visualized as a network in form of nodes and links. The description of datasets is shown in Table 1.

Table 1. Description of telecom datasets

Dataset source	Nodes	Links/attributes
Twitter (S_{1T})	4	40,449
FaceBook (S_{1F})	4	64,567
TRAI (T_1)	4	8

The methodology adopted for telecom network analysis is shown in Fig. 1.

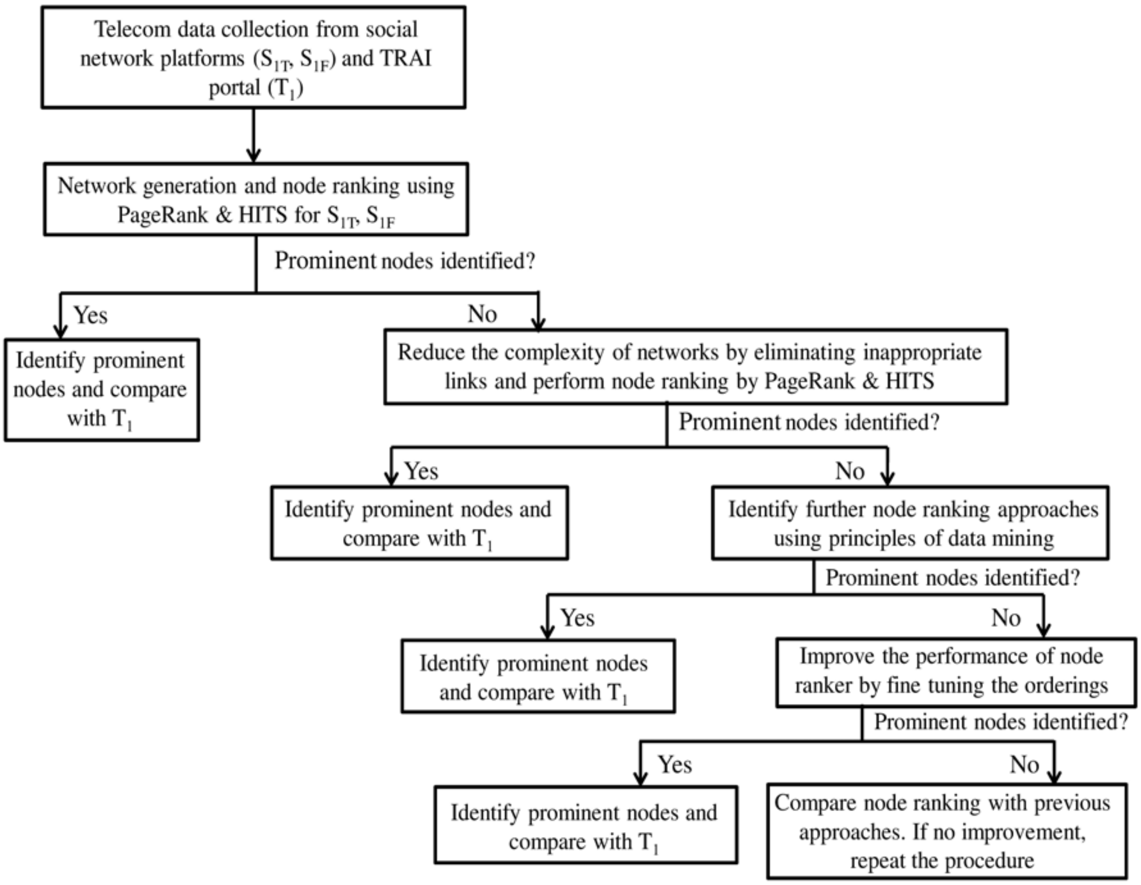
Telecom data collection from social network platforms (S_{1T}, S_{1F}) and TRAI portal (T_1)
Network generation and node ranking using PageRank & HITS for S_{1T}, S_{1F}
Prominent nodes identified?
Yes
No
Identify prominent nodes and compare with T_1
Reduce the complexity of networks by eliminating inappropriate links and perform node ranking by PageRank & HITS
Prominent nodes identified?
Yes
No
Identify prominent nodes and compare with T_1
Identify further node ranking approaches using principles of data mining
Prominent nodes identified?
Yes
No
Identify prominent nodes and compare with T_1
Improve the performance of node ranker by fine tuning the orderings
Prominent nodes identified?
Yes
No
Identify prominent nodes and compare with T_1
Compare node ranking with previous approaches. If no improvement, repeat the procedure

Fig. 1. Methodology adopted for telecom network analysis.

From these datasets, network structure is generated and visualized. The orientations of Twitter data is visualized from Gephi 0.9.2 at different sizes as Fig. 2. Similar orientations are also observed for Facebook data.

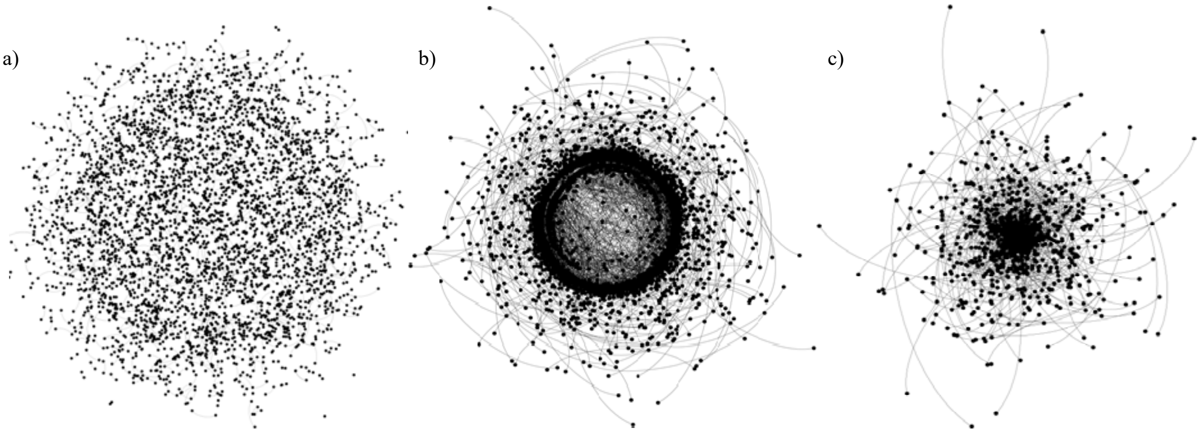

Fig. 2. Twitter network orientations. a) Initial data uploaded from twitter; b) Data visualized after calculating degree of certain nodes; c) Data visualized when clusters are being formed

Followed by data visualization, different centrality metrics are computed to estimate connectivity among nodes in networks. The calculations are tabulated in Table 2.

Table 2. Centrality metrics for social network data

Dataset	Centrality metrics	Value	Dataset	Centrality metrics	Value
S_{1T}	Degree centrality	22.34	S_{1F}	Degree centrality	26.72
S_{1T}	Clustering coefficient	0.652	S_{1F}	Clustering coefficient	0.693
S_{1T}	Eigenvector centrality	0.569	S_{1F}	Eigenvector centrality	0.604
S_{1T}	Cosine metric	0.228	S_{1F}	Cosine metric	0.382
S_{1T}	Jaccard metric	0.347	S_{1F}	Jaccard metric	0.371

Initial node ranking and link reduction

Once network structures are generated from social network datasets, node ranking is performed using PageRank and HITS to identify prominent nodes in the networks. The algorithms are executed in R programming language. Results derived from these algorithms are tabulated in Table 3. Here, the nodes are labeled based on their demarcations denoting different telecom operators.

Table 3. Results from PageRank and HITS algorithms on social network data prior to feature selection

Dataset	Algorithm	PageRank score		Authority score from HITS		Hub score from HITS	
S_{1T}	PageRank	n_1	0.022	-		-	
		n_2	0.024				
		n_3	0.021				
		n_4	0.020				
S_{1T}	HITS	-		n_1	0.019	n_1	0.033
				n_2	0.011	n_2	0.032
				n_3	0.016	n_3	0.029
				n_4	0.010	n_4	0.018
S_{1F}	PageRank	n_1	0.035	-		-	
		n_2	0.047				
		n_3	0.034				
		n_4	0.031				
S_{1F}	HITS			n_1	0.048	n_1	0.038
				n_2	0.035	n_2	0.040
				n_3	0.039	n_3	0.039
				n_4	0.041	n_4	0.036

As observed from the table, both algorithms fail to capture inherent topological differences among the nodes as it assigns analogous ranks to the nodes. Henceforth, link elimination techniques are to be implemented for reducing complexity of networks. However, connectivity of the networks is to be maintained prior and post link preprocessing operations. In this perspective, multiple links and loops are eliminated from the networks. Furthermore, boruta algorithm is chosen to perform link reduction owing to its capabilities in choosing relevant attributes from data. This algorithm chooses best features from data based on principles of random forest resulting in high scoring link attributes as output. Implementation of the algorithm in iterative mode on social network datasets results in simplified network with essential links. Results from this procedure are reflected in Table 4.

Table 4. Prominent connections derived after link reduction

Dataset	Nodes	Link features after removal of multiple edges and loop	Significant feature links identified by boruta	Mean importance score of top link feature	Iterations performed by boruta to reach convergence
S_{1T}	4	61, 318	402	145.32	1996
S_{1F}	4	38,954	361	192.74	2449
T_1	4	Not applicable	8	49.63	63

3.2. *Node ranking using ensemble approach*

Networks derived from previous step are used as inputs for node ranking procedures henceforth. These datasets are subjected to PageRank and HITS algorithms similar to previous iterations for detecting ordering of nodes. Differences in rankings are eminent compared to ranking on initial networks by the same algorithms. However, these algorithms tend to ignore relevant topological differences amongst the nodes resulting in analogous rankings. Results derived from these rankings are reflected in Table 5.

Table 5. Node ranking after removal of irrelevant link features

Dataset	Algorithm	PageRank score		Authority score from HITS		Hub score from HITS	
S_{1T}	PageRank	n_1	0.624	-		-	
		n_2	0.627				
		n_3	0.633				
		n_4	0.619				
S_{1T}	HITS	-		n_1	0.590	n_1	0.614
				n_2	0.601	n_2	0.623
				n_3	0.596	n_3	0.621
				n_4	0.593	n_4	0.617
S_{1F}	PageRank	n_1	0.600	-		-	
		n_2	0.601				
		n_3	0.604				
		n_4	0.610				
S_{1F}	HITS	-		n_1	0.640	n_1	0.621
				n_2	0.642	n_2	0.625
				n_3	0.639	n_3	0.618
				n_4	0.644	n_4	0.639

Both the rankers tabulate scores with minimal differences amongst all the nodes indicating probability of bias in orderings. Due to these marginal differences it is challenging to pin down a node as prominent. Henceforth, other ranking approaches needs to be identified for these networks. In this direction, SVM learner is adopted owing to its supremacy in node ranking applications [Lee (2014)]. Node ranking is performed using SVM learner on both S_{1T} and S_{1F} networks. The weight function w is modelled until converge is reached in the rankings. The scores derived from SVM are displayed in Table 6.

Table 6. Rankings derived from SVM

Dataset employed	No. of nodes in the network	SVM node ranking	
S_{1T}	4	n_1	0.734
		n_2	0.781
		n_3	0.729
		n_4	0.684
S_{1F}	4	n_1	0.841
		n_2	0.873
		n_3	0.766
		n_4	0.837

As observed from Table 6, SVM ranks the nodes diversely in both the datasets compared to previous algorithms. Despite better ranking compared to previous approaches, SVM also results in certain nodes having marginal differences. In case of S_{1T}, two nodes (n_1, n_3) receive analogous ranks while nodes (n_1, n_4) receive equivalent ordering in case of S_{1F}. Henceforth, rankings from SVM too cannot be considered for identifying prominent nodes. However, these rankings can be upgraded to distinguish equivalently ordered nodes.

In this direction, it is preferred to develop a ranker which can improve orderings of the pre-ranked SVM. RNN suits this requirement as it a variant of neural networks that feeds the output of previous layers as input to its current iteration. In this scenario, RNN can be tested by feeding rankings from SVM learner as input. This ensemble architecture (SVMRNNRank) must ensure that nodes are ranked based on their significances in the networks. Implementation of this approach is shown as Algorithm 1.

Algorithm 1: Node ranking by SVMRNNRank

```
Input: Node rankings from SVM for $S_{1T}$ & $S_{1F}$
Output: Node ranking from SVMRNNRank
1. Configure a RNN by feeding node rankings from SVM for $S_{1T}$
2. For each node N in $S_{1T}$, compute the salience score S(N)
3. Generate hierarchical representation of the nodes based on S(N)
4. Recompute S(N) by updating the weight and error functions in backpropogation mode
5. Iterate step (4) until convergence is reached
6. Terminate the computation and analyze the node rankings derived
7. Reconfigure the architecture by repeating steps 1-6 for $S_{1F}$ network
```

Nodes ranked from this approach are displayed in Tables 7and 8 respectively for S_{1T} and S_{1F} networks.

Table 7. Node rankings on S_{1T} by ensemble SVMRNNRank

Dataset	Node ranking		Error estimate	No. of iterations
S_{1T}	n_1	0.803	0.231	100
	n_2	0.801		
	n_3	0.793		
	n_4	0.384		
S_{1T}	n_1	0.805	0.228	200
	n_2	0.794		
	n_3	0.785		
	n_4	0.393		
S_{1T}	n_1	0.814	0.211	300
	n_2	0.808		
	n_3	0.778		
	n_4	0.391		
S_{1T}	n_1	0.811	0.202	400
	n_2	0.794		
	n_3	0.773		
	n_4	0.390		
S_{1T}	n_1	0.817	0.119	500
	n_2	0.795		
	n_3	0.772		
	n_4	0.390		
S_{1T}	n_1	0.816	0.119	600
	n_2	0.795		
	n_3	0.771		
	n_4	0.391		
S_{1T}	n_1	0.815	0.119	700
	n_2	0.794		
	n_3	0.771		
	n_4	0.390		

Table 8. Node rankings on S_{1F} by ensemble SVMRNNRank

Dataset	Node ranking		Error estimate	No. of iterations
S_{1F}	n_1	0.799	0.329	100
	n_2	0.763		
	n_3	0.692		
	n_4	0.444		
S_{1F}	n_1	0.801	0.334	200
	n_2	0.769		
	n_3	0.692		
	n_4	0.446		
S_{1F}	n_1	0.804	0.331	300
	n_2	0.768		
	n_3	0.691		
	n_4	0.445		
S_{1F}	n_1	0.806	0.329	400
	n_2	0.766		
	n_3	0.691		
	n_4	0.445		
S_{1F}	n_1	0.805	0.328	500
	n_2	0.764		
	n_3	0.690		
	n_4	0.443		
S_{1F}	n_1	0.809	0.320	600
	n_2	0.763		
	n_3	0.689		
	n_4	0.441		
S_{1F}	n_1	0.808	0.321	700
	n_2	0.765		
	n_3	0.688		
	n_4	0.443		
S_{1F}	n_1	0.802	0.317	800
	n_2	0.769		
	n_3	0.684		
	n_4	0.449		
S_{1F}	n_1	0.803	0.316	900
	n_2	0.769		
	n_3	0.683		
	n_4	0.443		
S_{1F}	n_1	0.804	0.316	1000
	n_2	0.766		
	n_3	0.682		
	n_4	0.444		

As observed from the tables, the ensemble approach ranks the nodes appropriately by highlighting distinctions among them. In case of S_{1T}, the algorithm ranks node (i.e. n_1) as significant based on its scores. Equivalently, the approach ranks same node (i.e. n_1) as prominent from S_{1F}. Marginal differences observed among multiple nodes from SVM ranker have been enhanced in these rankings. Validation of this approach is performed using statistical metrics.

3.3. *Performance analysis among node rankers*

It is essential to validate ranking efficacy of above mentioned approaches to ensure that nodes are not ordered randomly. In this context, precision and *MAP* values are computed for all four ranking algorithms i.e. PageRank, HITS, SVM, SVMRNNRank. The estimates are presented for node n_1 in Table 9.

Table 9. Performance analysis of node rankers with reference to node n_1

Ranking algorithm	**Precision for n_1 (S_{1T})**	***MAP* for n_1 (S_{1T})**	**Ranking order for S_{1T} network**	**Precision for n_1 (S_{1F})**	***MAP* for n_1 (S_{1F})**	**Ranking order for S_{1F} network**
PageRank	0.5	0.433	n_3, n_2, n_1, n_4	0.25	0.292	n_4, n_3, n_2, n_1
HITS	0.25	0.228	n_2, n_3, n_4, n_1	0.5	0.487	n_4, n_2, n_1, n_3
SVM	0.75	0.693	n_2, n_1, n_3, n_4	0.72	0.769	n_2, n_1, n_4, n_3
SVMRNNRank	0.94	0.895	n_1, n_2, n_3, n_4	0.91	0.883	n_1, n_2, n_3, n_4

As observed from table, it is evident that ensemble approach surpasses other techniques with better precision and *MAP* values for both network datasets. These estimates also indicate that SVMRNNRank ranks networks analogously maintaining local and global connectivity. Furthermore, ranking of nodes using this approach helps in identifying prominent telecom players from networks. Due to equivalent ordering from both networks, prominent telecom players are identified as Bharti Airtel (n_1) and Reliance Jio (n_2).

3.4. *Correlating S_1 and T_1 metrics*

Players identified from S_1 needs to validated with authentic performance parameters in T_1. Data extracted between timeframe November 2019 to March 2020 is analyzed to detect whether audience sentiments are in par with telecom standards. Eight standard parameters are considered for comparison namely average throughput for packet data (Kbps), call drop rate, minimum data download speed (Kbps), latency, service activation, successful data transmission for download, successful data transmission for upload and PDP context activation success rate. Trends from social network data (S_{1T}, S_{1F}) equivalent to the eight parameters are examined to detect customer sentiments. As a sample, parameters derived from state Karnataka are chosen for S_1 and T_1. They are analyzed to validate correlation among customer's opinions and standard metrics. The plot so derived from both datasets for different telecom parameters is shown in Fig. 3.

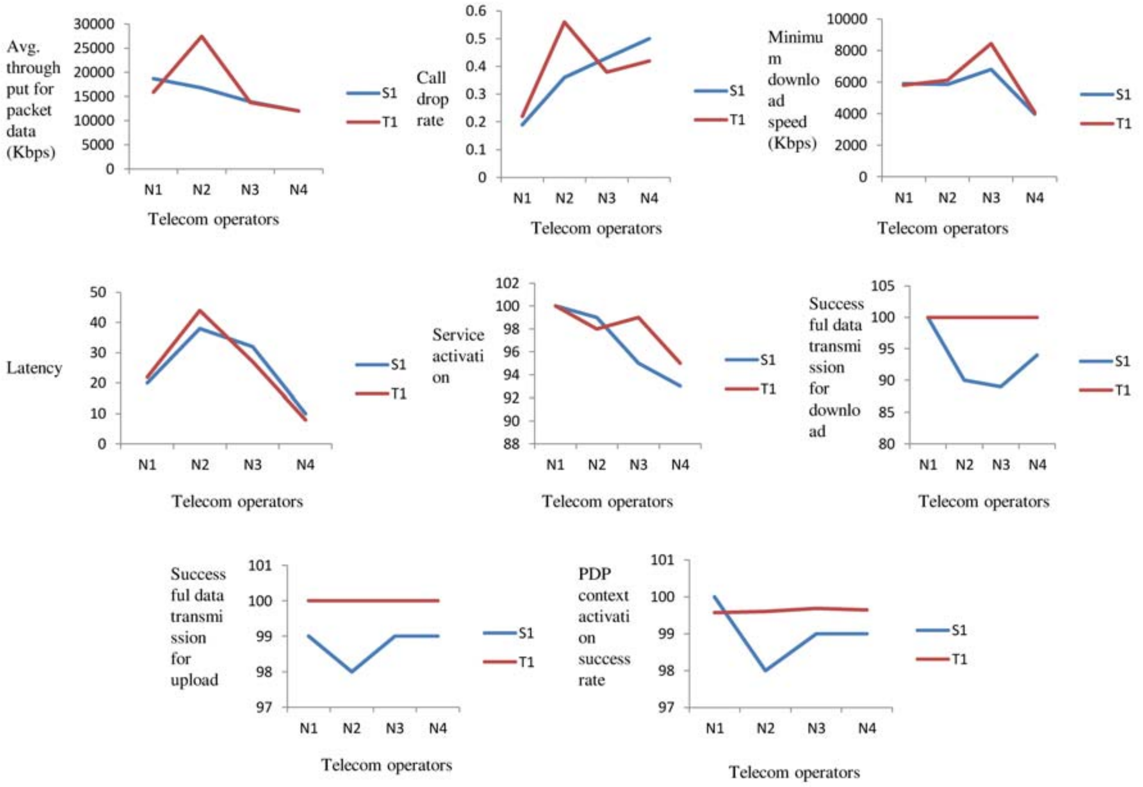

Fig. 3. Comparison of T_1 performance metrics with S_1 emotions

As observed from the plot, most parameters are distributed uniformly among operators n_1 and n_4 (i.e. Bharti Airtel and BSNL) with Airtel having best and BSNL having worst performances. However, there are variations among n_2 and n_3 (i.e. Reliance Jio and VodafoneIdea) in certain parameters (for instance, successful data transmission for upload, PDP context activation success rate) for identifying the next best service provider after Airtel. These findings are particular to the state of data collection (i.e. Karnataka) which may vary in other states owing to different estimates of the parameters.

Henceforth, this sample study reveals that there is positive correlation between customer emotions and standard parameters in ascertaining a superlative wireless telecom service provider in India.

4. Conclusion

Current study is an attempt to identify prominent service providers in telecommunication sector. In this direction, a network based approach is implemented to rank the telecom operators (i.e. nodes) using node ranking algorithms. A novel methodology is designed based on SVM and RNN that ranks the nodes by analyzing the global and local interconnections amongst nodes. This technique is implemented on real time customer sentiments extracted from social networking platforms, Twitter and FaceBook. Results deduced indicate better performance by the ensemble approach in ranking nodes. Furthermore, these rankings are used to detect key players amongst networks. It is revealed that Bharti Airtel has constructive influence while BSNL has slightest impression on social network users. Comparison of these outcomes with TRAI performance metrics indicates strong correlation with customer sentiments towards telecom operators. These results could be further improvised on larger timeframe and bigger samples to design effective strategies for customer retention and engagement. These approaches will in turn reduce churn rates of service providers as they may decode emotions of customers towards their service operators.

Acknowledgement

The authors are grateful to Visveshvaraya Technological University, Belgaum for giving an opportunity to conduct research and KSTePS, Government of Karnataka, for providing financial assistance to this work.

References

[1] Andresen, Philip. (2017): Acquiring customers through Social Customer Relationship Management An explorative case study within the telecom industry, pp. 1-37, 2017. Available at: https://ltu.diva-portal.org/smash/get/diva2:1115381/FULLTEXT01.pdf. Retrieved on 22nd April 2020.

[2] Barbera, P. (2017): Package 'Rfacebook', pp. 1-25. Available at: https://cran.r-project.org/web/packages/Rfacebook/Rfacebook.pdf. Retrieved on 5th November 2019.

[3] Bastian, M.; Heymann S.; Jacomy M. (2009): Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media, pp. 361-362.

[4] Cao, Ziqiang.; Wei, Furu.; Dong, Li.; Li, Sujian.; Zhou, Ming. (2015): Ranking with Recursive Neural Networks and Its Application to Multi-Document Summarization, Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Texas, USA, pp. 2153-2159.

[5] Chaparro, Cameron.; Eberle, William. (2015): Detecting Anomalies in Mobile Telecommunication Networks Using a Graph Based Approach, Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference, pp. 410- 415.

[6] Dahl, George E.; Yu, Dong.; Deng, Li.; Acero, Alex. (2012): Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition, IEEE Transactions on Audio, Speech, and Language Processing, 20(1), pp. 30-42.

[7] Deo, Anand. (2017): Telecom Industry in India: Evolution, Current Challenges & Future Road Map, Indira Management Review, 11(1), pp. 92-105.

[8] Feinerer, I. (2019): Package 'tm', pp. 1-64. Available at: https://cran.r-project.org/web/packages/tm/tm.pdf. Retrieved on 30th March 2020.

[9] Getoor, Lise.; Diehl, Christopher P. (2005): Link Mining: a survey, ACM SIGKDD Explorations Newsletter, 7(2), pp. 3-12.

[10] Gómez, Daniel.; Figueira, José Rui.; Eusébio, Augusto. (2013): Modeling centrality measures in social network analysis using bi-criteria network flow optimization problems. European Journal of Operational Research, 226(2), pp. 354-365.

[11] Gupta, Subhashish.; Tyagi, Kalpana.; Upadhyay, Rajkumar. (2018): Twilight of Voice, Dawn of Data: The Future of Telecommunications in India, IIMB-Working Paper No-563, pp. 1-49. Available at: https://www.iimb.ac.in/sites/default/files/2018-06/Twilight%20of%20Voice%2C%20Dawn%20of%20Data%20The%20Future%20of%20Telecommunications%20in%20India.pdf. Retrieved on 2nd April 2020

[12] Güven, Çiçek.; Atzmueller, Martin. (2019): Applying Answer Set Programming for Knowledge-Based Link Prediction on Social Interaction Networks, Front. Big Data, 2, pp. 1-15.

[13] Herbrich, Ralf.; Graepel, Thore.; Obermayer, Klaus (2000): Support Vector Learning for Ordinal Regression, MIT Press, Cambridge, pp. 1-6.

[14] IAMAI (2019): https://cms.iamai.in/Content/ResearchPapers/d3654bcc-002f-4fc7-ab39-e1fbeb00005d.pdf. Retrieved on 5th April 2020.

[15] J, Hidayati.; L, Ginting.; H, Nasution. (2018): Customer behaviour for telecommunication service provider, Journal of Physics: Conference Series, 1116, pp. 022015.

[16] Kearney, MW. (2020): Package 'rtweet', pp. 1-83. Available at: https://cran.r-project.org/web/packages/rtweet/rtweet.pdf. Retrieved on 3rd November 2019.

[17] Key Highlights of Economic Survey 2018-19 https://www.ibef.org/download/Key_Highlights_of_Economic_Survey_2018-19.pdf, pp. 1-9. Retrieved on 10th April 2020.

[18] Kim, YongSeog.; Street, Nick W.; Menczer, Filippo. (2003): Feature selection in data mining. Data mining: opportunities and challenges, pp. 80-105.

[19] Kleinberg, J. (1999): Authoritative sources in a hyperlinked environment. Journal of ACM, 46(5), pp. 604–632.

[20] Kursa, Miron Bartosz. (2018): Package 'Boruta' Wrapper Algorithm for All Relevant Feature Selection. https://cran.r-project.org/web/packages/Boruta/Boruta.pdf

[21] Lee, Ching-Pei.; Lin, Chih-Jen. (2014): Large-scale Linear RankSVM, Neural Computation, 26(4), pp. 781-817.

[22] Lim, Marcus.; Abdullah, Azween.; Jhanjhi, NZ.; Supramanium, Mahadevan. (2019): Hidden Link Prediction in Criminal Networks Using the Deep Reinforcement Learning Technique, Computers, 8(1), pp. 1-8.

[23] Liu, Guangyi.; Jiang, Dajie. (2016): 5G: Vision and Requirements for Mobile Communication System towards Year 2020, Chinese Journal of Engineering, pp. 1-8.

[24] Mariani, Manuel Sebastian.; Medo, Matúš.; Zhang, Yi-Cheng (2015): Ranking nodes in growing networks: When PageRank fails. Scientific Reports, 5(16181), pp. 1-10.

[25] Mishra, Sushruta.; Mishra, Brojo Kishore.; Tripathy, Hrudaya Kumar.; Mishra, Monalisa.; Panda, Bijayalaxmi. (2018): Use of Social Network Analysis in Telecommunication Domain, Modern Technologies for Big Data Classification and Clustering, pp. 152-178.

[26] Page, L.; Brin, S.; Motwani, R.; Winograd, T. (1998): The PageRank citation ranking: bringing order to the web. Stanford University, Computer Science Department Technical Report.

[27] Pinheiro, Carlos Andre Reis. (2011): Social Network Analysis in Telecommunications, Wiley Publishers, pp. 1-304, 2011.

[28] R, Insani.; H L, Soemitro. (2016): Data mining for marketing in telecommunication industry, 2016 IEEE Region 10 Symposium (TENSYMP), Bali, pp. 179-183.

[29] Reinsel, David.; Gantz, John.; Rydning, John. (2018): The Digitization of the World From Edge to Core. Available at: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf. Retrieved on 30th March 2020.

[30] S, Rajeshwari.; Srinivasulu, Yarlagadda.; Thiyagarajan, S. (2017): Relationship among Service Quality, Customer Satisfaction and Customer Loyalty: With Special Reference to Wireline Telecom Sector (DSL Service), Global Business Review, 18(4), pp. 1041-1058.

[31] Telecom Regulatory Authority of India (TRAI): https://www.trai.gov.in/release-publication/reports/wireless-data-reports. Retrieved on 3rd November 2019.

[32] Yadav, Rajesh K.; Dabhade, NIshant. (2019): Dynamics for Satisfaction with Telecom Services: A Pragmatic Investigation on Customers of Bhopal, Madhya Pradesh, Indian Journal of Marketing, 49(5), pp. 49-59.