Sustainable Computing: Informatics and Systems 24 (2019) 100351

Contents lists available at ScienceDirect

Sustainable Computing: Informatics and Systems

journal homepage: www.elsevier.com/locate/suscom

Social network sustainability for transport planning with complex interconnections

Somya Jain*, Adwitiya Sinha

Jaypee Institute of Information and Technology, Noida, India

ARTICLE INFO

Article history:
Received 30 January 2019
Received in revised form 22 May 2019
Accepted 20 September 2019
Available online 26 September 2019

Keywords:
Sustainable computing
Social network
Transport planning
Network bottleneck
Checkpoints
Large-scale graph analytics

ABSTRACT

Social network analysis serves as sustainable mechanism to examine large-scale complex social connections, with heterogeneity and interdependencies posing as major challenges. In our research, a novel approach is developed to efficaciously discover critical nodes, designated as network bottlenecks. The bottleneck is considered crucial in propagating the flow of information in the network. This is further extended to extraction of relative checkpoint(s) that acts as probable sources of major inflows towards the respective bottleneck. These set of checkpoints can be considered for prior surveillance resulting the control of information outbursts towards bottleneck node. Viable domains for applicability of our proposed methodology include, road traffic monitoring, extremist content tracking, fake news inspection, uncloaking online terrorist movements, etc. For our experimentation, we have focussed on transport planning application to identify traffic hotspot regions and relative set of nodes acting as checkpoints. These checkpoints can serve as monitoring stations for controlling the traffic hence improving sustainable mobility over roads. Moreover, air purifying machines can also be deployed, hence facilitating improved air quality.

1. Introduction

The inception of social networks since the mid of 20th century have resulted in formation of online social circles which eventually became an integral part of analysing the virtual interactions and dynamic complexities. The complex social structures largely simulate the realistic human behaviours, which can be further represented using mathematical and probabilistic modelling. Discrete structure such as graphs provides useful abstractions to captivate facets for understanding the underlying structure in social networks. A wide variety of applications, including viral marketing [1], intelligent transport system [2], cyber-crime detection [3,4], business intelligence [5], law enforcement [6], sustainability in biological life [7], financial risks in payment networks [8], etc., illustrates the potential of versatile applicability of social networking across diverse disciplines. The analytical strategies of social networks have been leveraged by engineers, mathematicians, biologist, geographers, physicists and scientists from various domains to investigate underlying complexities in respective networked frameworks. Network structure is composed of nodes with links connecting them, thereby exhibiting interaction and association. Social Network Analysis (SNA) not only assists researchers in evaluating the underlying framework of directed or undirected networks, but also extensively used for empirical analysis of weighted, signed and bipartite networks [9]. Exploration of complex networks and associated behaviours can be analysed using phenomenal methods and metrics of social networking, including homophily, centrality, clustering-coefficient, cohesion, density, propinquity, community detection, etc. [10–19].

Among sustainable categories of large-scale network-oriented problems [20], Intelligent Transportation System (ITS) is considered as one of the most conceivable areas of research that provides realistic solution for smart cities. The major protagonist in ITS is developing contingency plan for confronting epidemic traffic outbreaks in metapopulated areas, which often leads to severe air pollution in the ecosystem. In order to employ intelligence trait with transportation system, it must be equipped with information fusion facilities, computational, sensing, and wireless communication technologies.

For instance, transportation networks i.e. a network of roads, streets, airlines, metros and railways are using pedagogy of SNA to scrutinize its complex structure. The driving factors of coupling SNA with ITS are to infer traffic conditions, optimized route selection, information sharing, situation awareness, traffic incident detection, activity-travel patterns and many more [2,21–23].

* Corresponding author.
E-mail address: somya.jain@jiit.ac.in (S. Jain).

https://doi.org/10.1016/j.suscom.2019.100351
2210-5379/

Our research illustrates innovated usage of centrality theory from social networking domain for implementing sustainable transportation. The key highlights of our research work are stated as follows:

- A utilitarian algorithm is developed by infusing statistical central tendency
- Case study is conducted with large-scale road-network of California, having complex intersections and dense connections
- Structural properties of the extracted network is analysed with degree distribution, diameter, density, average path length, etc.
- The novel approach initiates with identification of most congested intersections, regarded as network bottlenecks
- Our proposed algorithm further assists in revealing the relatively prominent traffic hotspots (or checkpoints) leading towards globally significant junction(s), i.e. bottlenecks
- Detection of checkpoints is coupled with sustainable solutions in disaster management and decision support framework

Rest of the paper is organized in the following manner: second section will give glimpse of the related work or literature of this field. Third section will discuss overall proposed methodology and algorithms adopted for the research work. In addition, applications are also mentioned in this section. Fourth section will cover the dataset collection details and the experimental results which are followed by the fifth section concluding the research work done in the paper.

2. Related work

This section is constructed to succinctly describe the background of spatial transport network employing incredible properties of social network structures and their analysis methodologies in recent past. One of the study where the researchers [24], explored Boston underground transportation network consisting of 124 nodes (stations) and 124 edges (connecting tunnel to station) to investigate the existence of small world phenomena (six degree separation). Their study was interpreted using local and global efficiency whose foundation is based upon the notion of homophily and weak ties in social network. Also, authors in [25], applied network properties such as average path length, clustering coefficient and average node degree was to infer the small world phenomena in two railway line networks of Boston and Vienna. Networks considered was composed of 124 and 76 nodes (stations) connected with 8 and 5 edges (train lines) respectively. Comparative analysis with random bipartite graphs was also demonstrated. However, 22 public transport networks formed by bus and tramways spatial networks whose size varies from 152 to 2881 nodes were studied in [26] through statistical features of networks. Apart from clustering coefficient and average path length, degree distribution and medial based centrality was also found out. Another study of [27] note where the authors analysed bus transport network of China's four eminent cities to discover the occurrence of power-law (scale-free property) or Poisson (randomness property) degree distribution in the network. The result shows the distribution of number of nodes (bus stations) in a bus route and the number of bus routes a node will join in a network. In [28], vulnerability analysis of the Shanghai subway network was done to check robustness of the network through highest betweenness scored node and removal of critical nodes. The scenario was tested and compared against random network. One of the study where the analysis of urban transportation network comprising five major cities of Hungarian to understand their transportation system organization was presented with some significant statistical parameters [29]. The type of networks used was directed and weighted upon which global parameters like diameter, eccentricity distribution, community structure, average path length, and some centrality measures were figured out.

Also the authors in [30] have rendered the usage of social media information for travel and traffic planning to improve the overall transportation system. It order to extract the social data, overview of text mining process was also given. In [31], perspectives of collaborating social network with vehicular adhoc networks were given for improving the overall route planning and recommender system. In addition, emergent applications of vehicular social networks was also surveyed and mentioned. In [21], comprehensive study of vehicular social network was given including its attributes, applications, gaps with online social networks, contributions, and research challenges. Dynamics of Israel transportation network comprised of 6000 nodes and 15,000 connecting edges was studied by the authors in [2]. The network inferred was directed upon which the concept of betweenness and group betweenness was employed for better traffic assignment in a network. In recent years researchers [32] have tried to use new centric measure such as mobility centrality measure based upon kinetic energy with the sole purpose of probing network flows. The underlying theme of centrality was examined over Greek interregional commuting network constituting 39 nodes and edges each. Linear regression backward elimination method and Pearson's correlation coefficient were the evaluating parameters for testing the proposed index.

Performance evaluation of centrality metrics were conducted in [33] with Singapore subway network, consisting of 89 stations (nodes) for identifying the critical nodes with the help of medial (betweenness) and radial (degree and closeness) based centralities. Three centralities by incorporation of two factors i.e. commuter flow and travel time delay were designed and their performance was compared with classical centralities. In [34], intervention of world-wide air transportation network structure was carried out and results predict that the structure exhibits small world phenomena and scale-free property. Degree and betweenness centrality was also calculated to find out the most central node but the conception behind it was proved wrong due to occurrence of multi community in a particular network. Moreover, 33 metro systems

Table 1
Network Specific General Symbols.

Symbol	Description
G	Undirected graph
$\vec{G}$	Directed graph
N	Set of nodes in graph G or $\vec{G}$
E	Set of edges in graph G or $\vec{G}$
n	Number of nodes in graph, $n = \|N\|$
e	Number of edges in graph, $e = \|E\|$
$l_{geo}(s, d)$	Geodesic distance from node s to d where $s, d \in N$
$l_{geo,i}(s, d)$	Geodesic distance from node s to d where $s, d \in N$ for each path i
$\delta(s, d)$	Set of shortest paths from node s to d
$\left\|\delta(s, d)\right\|$	Total number of shortest paths from node s to d
$\delta_x(s, d)$	Set of shortest paths in which node x lies from s to d where $x \in N$
$\left\|\delta_x(s, d)\right\|$	Total number of shortest paths in which node x lies from s to d where $x \in N$
$B_x(G)$	Betweenness centrality for node $x \in N$
ψ	Parameter to extract single node or list of nodes bearing higher chances of being congested.
V	Array of betweenness centrality of each node
W	selective array of top ψ bottlenecks
ζ	Parameter to extract single checkpoint or a list of cascading checkpoints for each bottleneck.
M	Set of bottleneck checkpoints
m_ζ	Individual cascading checkpoint with respect to bottleneck
$C_x\left(G\|\vec{G}\right)$	Cascading checkpoint score for node $x \in N$
$\left\|\delta_x(s, y)\right\|$	Total number of shortest paths in which node x lies from s to y where $x \in N$
$\left\|\delta(s, y)\right\|$	Total number of shortest paths from node s to y

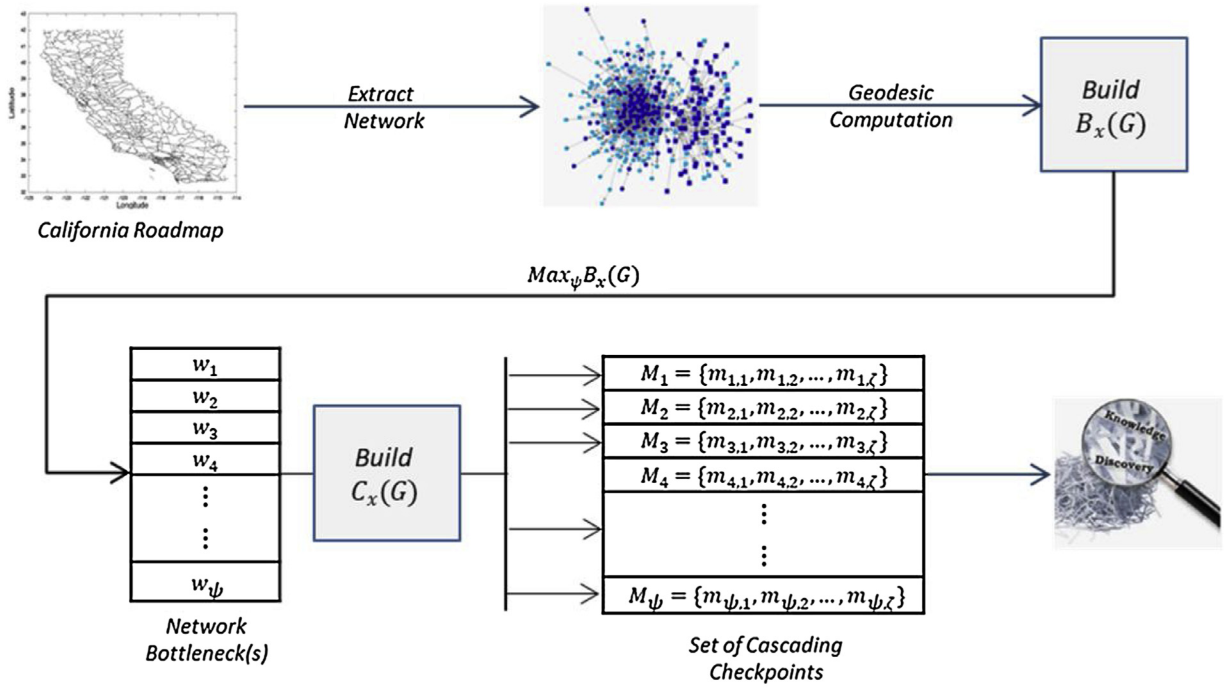


Fig. 1. Proposed Methodology Framework.

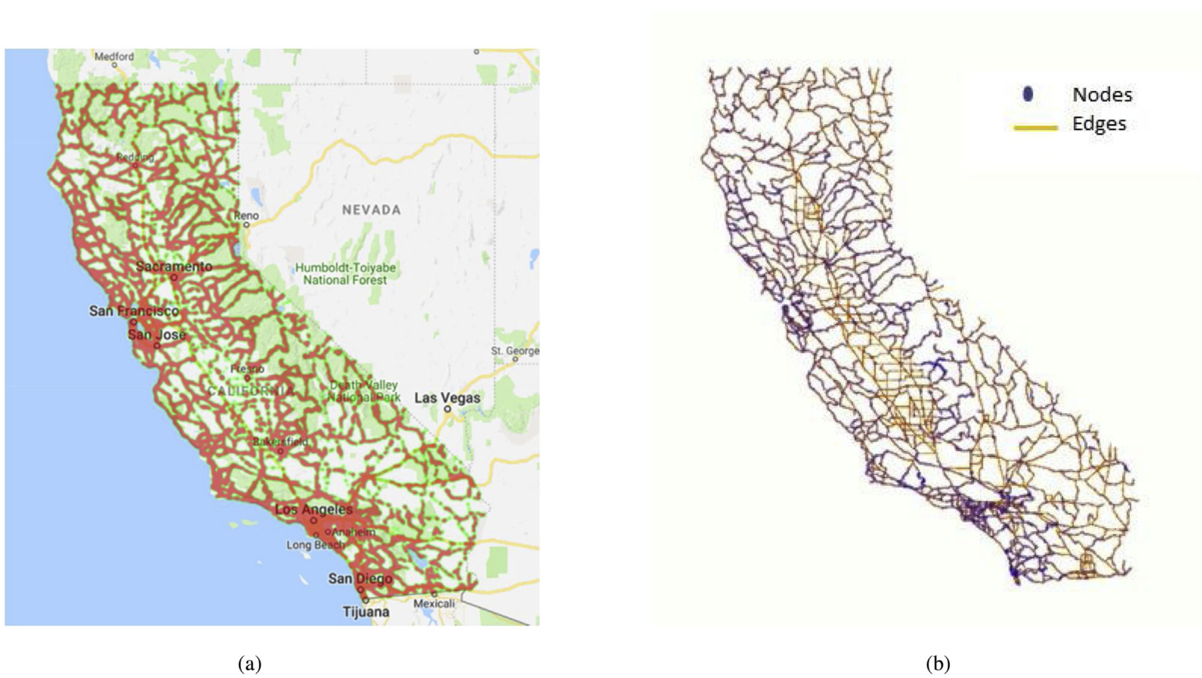


Fig. 2. Visualization of Road Map and Extracted California Transportation Network.

were taken in to account for experimental purpose in [35] and found out that most of the network's degree distribution follows power law and having hubs (some nodes have very large degree which are connected to nodes possess low degree). In addition to it, small world phenomena was also revealed in the network which was characterized by shortest paths and clustering coefficient. In [22], unified approach of Kalman filter model with social information was developed to predict bus arrival time for effective traffic management. Twitter social network was used to extract the information related to traffic from hastags and tweets posted by users.

Another research where the authors use social information was done in [36] where an exhaustive study was presented describing how the use of twitter social information and artificial neural network to predict traffic congestion in real-time and hence providing a valuable source for traffic analysis. In [23], Sina Weibo social platform of China was explored to efficiently extract information exchanged over it regarding traffic with the help of natural language processing techniques and human computer interaction. The overall procedure was done to provide a vital source to people for better travel planning. Intervention of core density and branch structure was observed in [37] for 14 largest cities subway net-

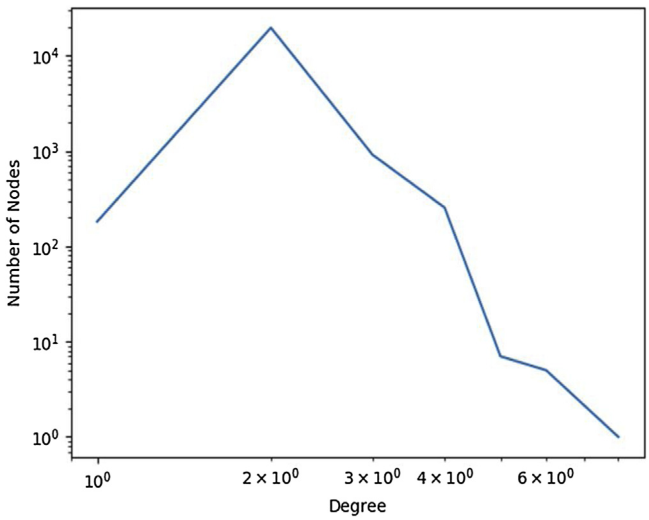


Fig. 3. Degree distribution graph of California Road Network.

works of world. The study includes the discussion of statistical features incurred and evolving structure of networks. In another study [38], data set of world largest subway system was investigated to determine robustness of these networks against random attacks through intentionally stalling the paths crossing targeted nodes, multiple attacks on high degree nodes and or by changing the size of giant component. The authors in [39], studies Guangzhou urban transportation network to find out prominent sites in a city using centrality index. Investigation over three case studies of Louisiana was presented in [40], where the researchers have analysed sub-urban street, urban street and urban transportation network of distinguished areas. Critical intersections were listed using Bonacich power, 2-step reach, betweenness and Eigen-vector social network analysis measures.

Traditionally, researchers have focussed on finding statistically significant trends in the various small scale transportation networks. Their work is considered as a comprehensive study because it delineates the foremost physiognomies of the underlying network. However, benchmark analytical techniques of social network were used only to identify hubs and congested intersection. The previous research lacked the investigation of ways to handle these critical nodes or to perceive it for further optimal deployment of traffic policies to control congestion. Our novel approach will reveal bottleneck nodes with high traffic probability that eventually poses greater chances of evolving into state of congestion. The proposed work builds model to detect the relative list of checkpoints for intelligent and sustainable transport planning. The proof of the overall proposed methodology is shown by induction on large scale road network.

Table 2
Structural properties of California road network.

Structural Properties	Values
Nodes	21048
Edges	21693
Average Degree	2.0613
Average Clustering Coefficient	7.126567844925884e − 05
Transitivity	0.0002486
Diameter	728
List of nodes in periphery	[′80′, ′20600′, ′20618′]
Average Path Length	270.5432481053562
Density	9.793740121947844e − 05

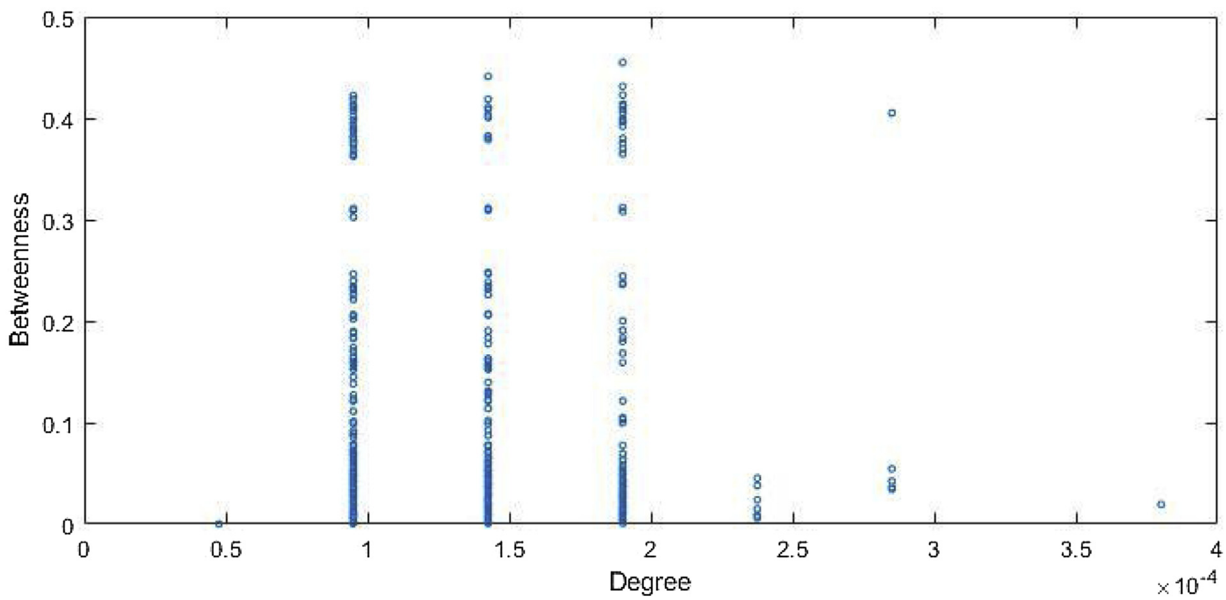
3. Proposed framework: significance & methodology

In this section, we present diverse motivations in terms of applications of proposed methodology that drive to deduce new perspectives. Next, we will illustrate the construction of proposed algorithm with a set of preliminaries which can be used to solve wider range of problems. This section will also illuminate the generalized form of supporting structure or the composition of our research flow.

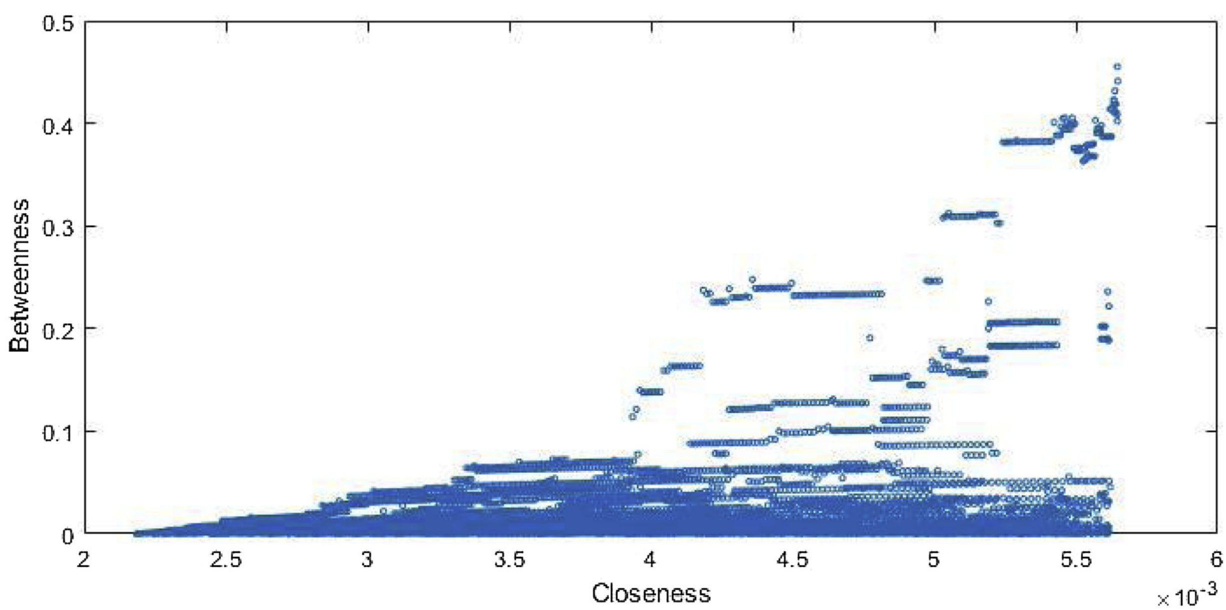
Some of the applications listed below which are identified to substantiate the significance of research:

- Intelligent Transportation System: Transportation has a strong influence on the spatial structure at the local, provincial and global levels. Since the impact of transport on contemporary society is very crucial, proposed algorithm helps to underline the importance of specific dimensions such as nodes, locations, networks and complex interactions. It will support to regulate the transport policies and planning by addressing the problem of rerouting flows through alternative routes and locations whenever there is congestion or any kind of natural and man-made disasters on road. Environmental sustainability can also be achieved by configuring anti-pollution devices at identified prominent junctions to purify the air.
- Information Dissemination: Nowadays, Social network platforms assist in transmission of information or rather spreading it to mass media such as in the form of advertisements, public announcements via social posts and blogs. To disseminate, one needs to rectify the proper communication channel or seed set to maximize the rate of spreading. By utilizing the approach, marketers and advertisers can find out dissemination patterns in the social and spatial network and hence can identify the minimum set of potential seeds for effective diffusion. These scattering seeds will cause in economic sustainability with limited budget.
- Cyber Criminology: Security is an integral part of society whether on online networks or offline networks. Today, identification of criminal groups or communities over social network is a primary concern. For instance, terrorist network can be analysed which are formed over online platforms and discovering the trend of their communication links will be helpful. Moreover, identifying leaders of the group and breaking their communication links will be great move towards security goal.
- Biological Networks: Methodologies of SNA is widely used for analysing biological networks for example, Gene structure, Protein network, Brain network etc. Identification of key connectors across the brain for information based mind control and controlling metabolites in underlying networks will help to examine elementary constitutes of them. Therefore, network profiling, extracting hidden properties and facts will add up to biological significance.
- Fake News Inspection: Among the core issues over online social platforms, virality of fake news over them has come at the forefront. Due to this spreading of valuable information is hindered and results in information loss. Therefore, inspection of root nodes and associated major communication junctions responsible for fake news propagation is necessary which aids in restrain from social damage. Situation awareness process will then be boosted if the prominent source of negative behaviour is interrupted.

Formalization of the approach is stated using some common notations and symbols described in Table 1. The overall process framework is presented in Fig. 1 starting from network extraction to knowledge discovery. Once the network is extracted, the goal is to find all the possible geodesics over it and build $B_x(G)$. We then begin building $C_x(G)$ by selecting top scored essential node by using



(a) Degree and Betweenness



(b) Closeness and Betweenness

Fig. 4. Visualization of Scatter Plots.

max function which leads to retrieve a set of cascading checkpoints. These points were obtained by iterative refining approach. Therefore, by locating these points, discovered knowledge will facilitate inferencing and triggers the cognitive process of decision making.

Let an undirected graph $G(N, E)$, where N is a set of nodes and E is a set of unordered pair of distinct nodes called undirected edges or links. $n = \left|N\right|$ denotes size of the given network modelled in form of graph. Similarly, a directed graph $\vec{G}$ is described as a pair $\overrightarrow{G}(N, \overrightarrow{E})$, where N is a set of nodes and E is a set of ordered pair of distinct nodes called directed edges having directivity associated with them. A path $i = \left\{N_0, E_1, N_1, E_2, N_2 \ldots\right\}$ is defined as alternating sequence or set of nodes and links, with no repeated edges or vertices. Geodesic distance between two distinct nodes i.e. $s, d \in N$ in a graph G or $\vec{G}$, denoted by $l_{geo}(s, d)$, refers to the minimum number of edges traversed to reach the destination node d from the source vertex s. Geodesic distance is designated as the shortest path connecting a pair of nodes. In case of undirected graph G, geodesic distance $l_{geo}(s, d)$, from source s to distance d, will be symmetric. However, for directed graph $\vec{G}$, geodesic distance will be asymmetric due to directivity of arcs, i.e. $l_{geo}(s, d) \neq l_{geo}(s, d)$.

In a connected graph, G or $\vec{G}$, it is possible that there exist more than one shortest path between same pair of nodes. Moreover, $l_{geo,i}(s, d)$ refers to the geodesic distance from node s to d for each path i to be further considered in computation of $\delta(s, d)$ and $\delta_x(s, d)$ respectively. The symbol $\delta(s, d)$ represents a set of shortest paths exist between nodes s to d. However, total number of shortest paths occurred between nodes s to d is given by $\left|\delta(s,d)\right|$. Set of shortest paths in which node x lies from s to d where $x \in N$ is denoted by $\delta_x(s,d)$ and $\left|\delta_x(s,d)\right|$ will give the number of shortest paths in which node x lies from s to d.

Our proposed approach focuses on discovering the node(s) highly probable to be congested, which may act as bottleneck(s) owing to its network connectivity. The problem is addressed by proposing a generic model that computes and extracts the list of node(s) having influential position in terms of connectivity and reachability. The betweenness centrality is one of the benchmark node-centric measures to identify such influential nodes. On the basis of control flowing through nodes in the interconnected networked structure, Bx(G) or Bx($\vec{G}$) represents the betweenness score of each node x in undirected or directed case. This node-specific score will encompass all the shortest routes or channels, which cross through the node. Therefore, if more routes passes through the node x, more congested it would become. Therefore, betweenness centrality is considered as a measure of congestion at any node and mathematically is defined as in Eqs. (1) and (2) [11]:

$$B_x(G) \longleftarrow \frac{2}{(\left|N\right|-1)(\left|N\right|-2)} \times \sum_{s,\,d\,|\,(s,\,d\,\neq x\,\in N)} \frac{\left|\delta_x(s,d)\right|}{\left|\delta(s,d)\right|} \tag{1}$$

$$B_x(\vec{G}) \longleftarrow \frac{1}{(\left|N\right|-1)(\left|N\right|-2)} \times \sum_{s,\,d\,|\,(s,\,d\,\neq x\,\in N)} \frac{\left|\delta_x(s,d)\right|}{\left|\delta(s,d)\right|} \tag{2}$$

In order to map the betweenness centrality over standard scaling range of [0, 1], $\binom{n-1}{2}$ is used as normalization factor for undirected case. In the directed case, the ordered pairs are twice as many as unordered pairs, hence the normalization factor becomes $2 \times \binom{n-1}{2}$. Betweenness centrality quantifies the number of times a node in a network acts as a bridge along the shortest paths between all pairs of nodes in the network. Our approach for finding out the nodes with high congestion probability is a two-phase methodology that involves finding network bottlenecks (proposed algorithm 1) and associated cascading checkpoints (proposed algorithm 2). Here, iterative refining strategy works by finding solution or cascading list of nodes leading to globally identified bottleneck by recursively breaking down a search space and fixing a destination node. Bottleneck node is identified nearly from a polynomial search space whereas associated checkpoints are identified from linear search space, excluding all-pair dependencies. The process begins with discovering the list of bottleneck points in the network, as described in proposed algorithm 1. The algorithm takes graph (G or $\vec{G}$) and a network parameter ψ that denotes an adjustable parameter to extract single node or list of nodes bearing higher chances of being congested. Resultantly, our model would return network bottleneck node(s), or simply bottlenecks(s). For instance, if $\psi = 1$, then the algorithm returns the most congested node amongst all nodes on the basis of betweenness centrality. This algorithm takes $O(e)$ time to determine the number of shortest paths from a given source to other vertices by using Breadth First Search (BFS) [11]. Hence, $\delta(s,d)$ can be computed with time complexity of $O(ne)$. Consequently, the runtime complexity of algorithm 1 becomes $O\left(n^2 e\right)$.

Proposed Algorithm 1: Network Bottleneck(s) Extraction

Input: $G(N,E)$ or $\vec{G}(N,\vec{E})$, ψ

Procedure:

$n = |N|$
$V := \emptyset$ // array of betweenness centrality of each node
for each node $x \in N$ set:
  $B_x(G) = 0$ // for undirected graph
  $B_x(\vec{G}) = 0$ // for directed graph
  $\delta(s,d) := \emptyset \quad \forall\, s,d \in N$
  $\delta_x(s,d) := \emptyset$
end-for
for each node x in N:
  for every pair of distinct nodes $s,d \in N$ and $s,d \neq x$:
    Build $\delta(s,d) := \bigcup_i l_{geo,i}(s,d)$ // union of all i geodesics
    Build $\delta_x(s,d) := \bigcup_{i|x\in i} l_{geo,i}(s,d)$ // union of all i geodesics having x
  end-for
  Compute $B_x(G)$ or $B_x(\vec{G})$
  $V = V \cup B_x$
end-for
Desc-Sort V
$W := \emptyset$ // selective array of top ψ bottlenecks
$W = \max_{\psi}(V)$

Output: Set of Network Bottlenecks, $W = \{w_1, w_2, \dots, w_\psi\}$

Next, the procedure of finding out the cascading checkpoints corresponding to each bottleneck node is instigated. Here, the proposed algorithm 2 takes set of network bottlenecks W, which was retrieved using proposed algorithm 1 as an input along with adjustable parameter ζ to extract single checkpoint or a list of cascading checkpoints for each bottleneck. For instance, if $\zeta = 1$, then the algorithm returns a single checkpoint node with respect to each network bottleneck whose cascading checkpoint score $C_x\left(G|\vec{G}\right)$ is maximum. Moreover, ζ parameter can be adjusted according to the vicinity of congestion range, which is normally restricted to few kilometres. This algorithm runs in $O(nm\zeta)$ time.

Proposed Algorithm 2: Cascading Checkpoints for Bottleneck(s)

Input: $G(N,E)$ or $\vec{G}(N,\vec{E})$, W, ζ

Procedure:

$W = \{w_1, w_2, \dots, w_\psi\}$ // bottleneck set
$\psi = |W|$
$i = 1, j = 1$
$M = \{M_1, M_2, \dots, M_\psi\}$ // checkpoint set
$M_i := \emptyset \ \forall_i$
$Z := \emptyset$
for bottleneck-node $w_i \in W$ do: // executed in a distributed manner for each bottleneck
  $N := N - \{w_i\}$
  $M_i := M_i \cup w_i$
  $y = w_i$
  while $j \le \zeta$ do:
    for each node $x \in N$:
      $C_x(G|\vec{G}) \leftarrow \frac{1}{(|N|-j-1)} \times \sum_{s,y\,|\,(s,y\neq x,\ s\neq y,\ s,x\in N)} \frac{|\delta_x(s,y)|}{|\delta(s,y)|}$
      $Z = Z \cup C_x$
    end-for
    Desc-Sort Z
    $y := \max(Z)$
    $N := N - y$
    $M_i := M_i \cup y$
    $j = j + 1$
  end-while
end-for

Output: Checkpoints Set for ψ Bottlenecks, $M = \{M_1, M_2, \dots, M_\psi\}$
Set of Cascading Checkpoints for i^{th} bottleneck, $M_i = \{m_{i,1}, m_{i,2}, \dots, m_{i,\zeta}\}$; $\forall i = 1 \dots \psi$

Therefore, for discovering the checkpoints or monitoring nodes to control congestion we can set threshold for how many units or nodes we needed to control long traffic snarls. Furthermore, by locating these nodes where heavy traffic crawls across them can be installed with air purifying machines, mounted air filters etc. to improve air quality. Also, public transport such as local buses crossing choked junctions can also be mounted with air purifiers. Furthermore, these nodes can also be used for advertisement hoardings i.e. marketing communication point because their location signifies high traffic vehicular movement and hence promotion gives lasting impact over there.

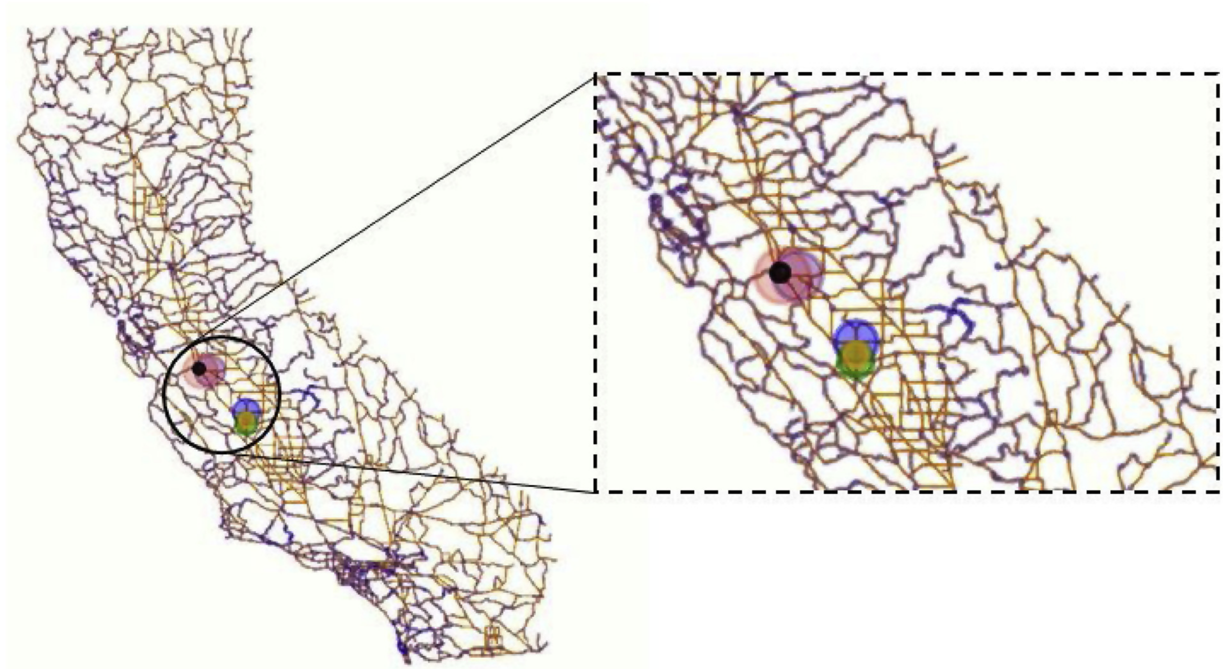


Fig. 5. Visualization of Top Network Bottlenecks.

4. Results & discussion

In this section the proposed model is demonstrated to find out some interesting facts over complex transportation network which is the base of whole empirical study. To assess the quantitative analysis top six selective bottlenecks are identified. We have considered dataset of California road network which is modelled using undirected graph comprises of 21,048 nodes and 21,693 edges [41]. Network measures and its structural properties were extracted and summarized in Table 2. From the data fetched, it was observed that road network formed was connected graph but do not follows the principle of small world phenomena. A small world phenomenon is characterized by average path length and average clustering coefficient. If average path length is very small as compared to clustering coefficient then graph is considered to exhibit the phenomena. Though a network is connected, it means one can reach from any node to every other node irrespective of how apart they are in a network. There is a set of three periphery nodes whose eccentricity is equal to diameter. Graphical visualization of the road network is shown in Fig. 2(a) and (b) respectively.

Network of nodes and edges extracted from the road map (nodes in purple and links in yellow) is shown in Fig. 2(b). For investigation we first analysed the structure of road network by calculating degree distribution of the nodes. Fig. 3 shows the degree distribution graph of underlying road network using log-log plot. We inferred that our network follows scale-free property because its degree distribution follows power law (no Poisson distribution) i.e. many nodes in a network with only few links but few nodes are with large number of links. These nodes are categorized as hubs. Therefore, by calculating local neighbourhood with respect to each node in a network only one node (node id: '14300′, local neighbourhood score: 0.00038) with eight connecting links is considered as a major hub. The presence of these hubs change the way of navigation on a road. Next observation about the structure of a network is that it is more centripetal rather than centrifugal network. Network is labelled as centripetal because here are some significant nodes having high connectivity than the others whereas in centrifugal networks there are no significant or important nodes on the basis of connectivity. Moreover, these networks have a grid-like layout or pattern. Closeness centrality is used to find out the extent of reachability by measuring which node is near to all segments in the road network along the shortest paths. It also captures the notion of accessibility of a place at the city scale.

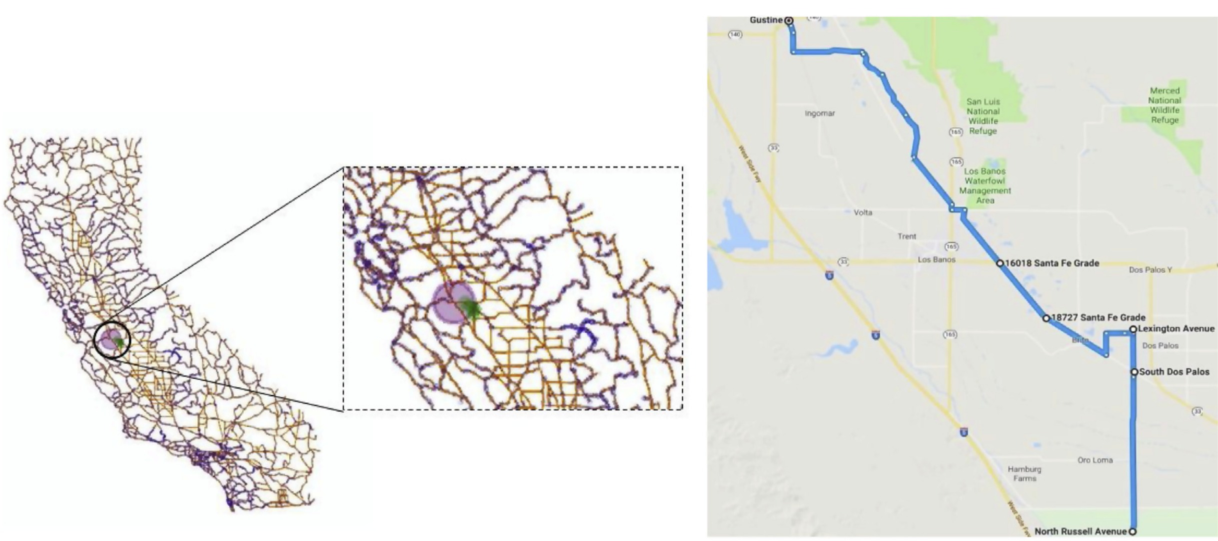
Therefore, the top scored nodes can be used as a facility provider location like petrol pumps, police station, restaurant, hospital etc. which aids in achieving a goal of sustainable development. After computing closeness score, node with a node id: '10214′ scores (0.00564) maximum.

Interestingly, it is discovered that out of 21, 048 nodes 182 nodes are border nodes depicting that these do not lies between any geodesic in a network. Hence, their $B_x(G)$ is zero. Scatter plot of all the betweenness values of each node in a network with respect to their degree and closeness values is shown in Fig. 4(a) and (b) respectively. Scatter plots visualizations aids in construe the appearance of linear relationship and the pattern of it over a straight line which are used to interpret trends in statistical data. Among other statistics, the output of scatter plot shows that degree and betweenness values are not correlated. However, closeness and betweenness values seem to be positively correlated depicting scatter points are accumulated below the straight line on the right.

For experiment purpose to identify potential nodes using proposed algorithm 1 and 2, value of adjustable parameters $\psi = 6$ and $\zeta = 5$ are considered. These values can be changed according to requirement. On running algorithm 1, we get the list of bottlenecks $W = \left\{10211,\ 10214,\ 11597,\ 11969,\ 11970,\ 10212\right\}$ which are highlighted in Fig. 5. For better understanding of the finding, visualization of network is highlighted with network bottlenecks in distinct colours, with first bottleneck in largest size. Next, with respect to each bottleneck we found out a list of five cascading checkpoints on the basis of checkpoint score. Table 3 shows the list of nodes with $B_x(G)$ and $C_x(G)$ score with corresponding identified places. Further, Fig. 6(a) and (b) provides the visualization of identified first bottleneck (purple) with associated checkpoints (green). In Fig. 6(b), specified view is visualized by highlighting a path with identified location (node(s)) names.

Table 3
network Bottlenecks & associated cascading checkpoints.

Set of Top Network Bottlenecks					Set of First Five Relative Checkpoints					
W	$B_x(G)$	Place	Longitude	Latitude	M_i		$C_x(G)$	Place	Longitude	Latitude
					$m_{1,1}$	10214	0.51887	16018,Santa Fe Grade	−120.787338	37.054626
					$m_{1,2}$	10516	0.44848	18727, Santa Fe Grade	−120.739197	37.01146
w_1 10211	0.45496	Gustine	−120.967216	37.054817	$m_{1,3}$	10514	0.48868	Lexington Ave, Dos Palos	−120.653328	36.97216
					$m_{1,4}$	10680	0.49082	South Dos Palos	−120.652969	36.964508
					$m_{1,5}$	10681	0.50147	Russell Ave, Firebaugh	−120.654579	36.898151
					$m_{2,1}$	10211	0.47816	Interstate-5, Gustine	−120.967216	37.054817
					$m_{2,2}$	10516	0.45006	18727, Santa Fe Grade	−120.739197	37.01146
w_2 10214	0.44119	Santa Fe Grade	−120.787338	37.054626	$m_{2,3}$	10514	0.48868	Lexington Ave, Dos Palos	−120.653328	36.97216
					$m_{2,4}$	10680	0.49082	South Dos Palos	−120.652969	36.964508
					$m_{2,5}$	10681	0.50147	Russell Ave, Firebaugh	−120.654579	36.898151
					$m_{3,1}$	11628	0.49276	Cantua Creek, CA 93608	−120.332779	36.426201
					$m_{3,2}$	11629	0.49352	Coalinga	−120.333519	36.445923
w_3 11597	0.43139	20860 Lassen Ave, Five Points	−120.10289	36.431538	$m_{3,3}$	11543	0.49488	CA-33, Cantua Creek	−120.395187	36.447189
					$m_{3,4}$	11550	0.49213	Cantua Creek, CA 93608	−120.404083	36.455925
					$m_{3,5}$	11551	0.49513	Cantua Creek, CA 93608	−120.426933	36.481041
w_4 11969	0.42289	Lassen Ave, Huron	−120.10289	36.257809	$m_{4,1}$	11970	0.52788	28937 Officer John Palacios Memorial Hwy, Five Points	−120.102463	36.314503
					$m_{4,2}$	11597	0.52847	20860 Lassen Ave, Five Points	−120.10289	36.431538
					$m_{4,3}$	11628	0.49281	Cantua Creek, CA 93608	−120.332779	36.426201
					$m_{4,4}$	11629	0.49357	Coalinga	−120.333519	36.445923
					$m_{4,5}$	11543	0.49493	CA-33, Cantua Creek	−120.395187	36.447189
					$m_{5,1}$	11597	0.52845	20860 Lassen Ave, Five Points	−120.10289	36.431538
w_5 11970	0.42209	28937 Officer John Palacios Memorial Hwy, Five Points,	−120.102463	36.314503	$m_{5,2}$	11628	0.49279	Cantua Creek, CA 93608	−120.332779	36.426201
					$m_{5,3}$	11629	0.49354	Coalinga	−120.333519	36.445923
					$m_{5,4}$	11543	0.49490	CA-33, Cantua Creek	−120.395187	36.447189
					$m_{5,5}$	11550	0.49215	Cantua Creek, CA 93608	−120.404083	36.455925
					$m_{6,1}$	10211	0.57205	Gustine	−120.967216	37.054817
					$m_{6,2}$	10214	0.51889	Santa Fe Grade	−120.787338	37.054626
w_6 10212	0.41876	Interstate-5	−120.993408	37.084145	$m_{6,3}$	10516	0.44850	18727, Santa Fe Grade	−120.739197	37.01146
					$m_{6,4}$	10514	0.48870	Lexington Ave, Dos Palos	−120.653328	36.97216
					$m_{6,5}$	10680	0.49084	South Palos	−120.652969	36.964508



(a) (b)

Fig. 6. Visualization of Most Congested Bottleneck and Checkpoints in a Network.

5. Conclusion

Social network analysis has great eminence in analysing and identifying informative patterns in a network. The network profiling when combined with knowledge extraction enhances its usefulness over variety of sustainable applications. Our approach investigated the significance of centrality theory to develop sustainable solution for efficient transport planning. The innovated technique assists in locating the crucial frontiers leading towards bottlenecks nodes, which are extremely likely to be congested. Novel algorithms were developed to identify cascading list of checkpoints in the large scale complex network. The experimentation is conducted over complex road network of California to illustrate and justify the practical applicability of our proposed algorithm. Simulation results highlight statistical and structural properties of the underlying network. Our experimental results provide insight onto an information space disclosing bottlenecks and checkpoints with graphical primitives.

References

[1] Leonidas Akritidis, Dimitrios Katsaros, Panayiotis Bozanis, Identifying the productive and influential bloggers in a community, IEEE Trans. Syst. Man Cybern. Part C 41 (5) (2011) 759–764.
[2] Rami Puzis, et al., Augmented betweenness centrality for environmentally aware traffic monitoring in transportation networks, J. Intell. Transp. Syst. Technol. Plan. Oper. 17 (1) (2013) 91–105.
[3] Mohammed Alzaabi, Kamal Taha, Thomas Anthony Martin, Cisri: a crime investigation system using the relative importance of information spreaders in networks depicting criminals communications, IEEE Trans. Inf. Forensics Secur. 10 (10) (2015) 2196–2211.
[4] Giles Oatley, Tom Crick, Measuring UK crime gangs: a social network problem, Soc. Netw. Anal. Min. 5 (1) (2015) 33.
[5] Jesús Bobadilla, et al., Recommender systems survey, Knowl. Based Syst. 46 (2013) 109–132.
[6] Nicola Lettieri, et al., A computational approach for the experimental study of EU case law: analysis and implementation, Soc. Netw. Anal. Min. 6 (1) (2016) 56.
[7] Jianxin Wang, et al., Identification of essential proteins based on edge clustering coefficient, IEEE/ACM Trans. Comput. Biol. Bioinf. (TCBB) 9 (4) (2012) 1070–1080.
[8] Ben Craig, Goetz Von Peter, Interbank tiering and money center banks, J. Financ. Intermediation 23 (3) (2014) 322–347.
[9] Evelien Otte, Ronald Rousseau, Social network analysis: a powerful strategy, also for the information sciences, J. Inf. Sci. 28 (6) (2002) 441–453.
[10] Linton C. Freeman, Centrality in social networks conceptual clarification, Soc. NetW. 1 (3) (1978) 215–239.
[11] Brandes Ulrik, A faster algorithm for betweenness centrality, J. Math. Sociol. 25 (2) (2001) 163–177.
[12] Miller McPherson, Lynn Smith-Lovin, James M. Cook, Birds of a feather: homophily in social networks, Annu. Rev. Sociol. 27 (1) (2001) 415–444.
[13] Paul W. Holland, Samuel Leinhardt, Transitivity in structural models of small groups, Comp. Group Stud. 2 (2) (1971) 107–124.
[14] Duncan J. Watts, Steven H. Strogatz, Collective dynamics of 'small-world'networks, nature 393 (1998) 440, 6684.
[15] James Moody, Douglas R. White, Structural cohesion and embeddedness: a hierarchical concept of social groups, Am. Sociol. Rev. (2003) 103–127.
[16] Charles Kadushin, Understanding Social Networks: Theories, Concepts, and Findings. OUP USA, 2012.
[17] Guandong Xu, Yanchun Zhang, Li. Lin, Web Mining and Social Networking: Techniques and Applications., 6, Springer Science & Business Media, 2010.
[18] R.éka Albert, Albert-L.ászló Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys. 74 (1) (2002) 47.
[19] Mark E.J. Newman, The structure and function of complex networks, Siam Rev. 45 (2) (2003) 167–256.
[20] I. Ahmad, S. Ranka, Handbook of Energy-Aware and Green Computing-Two Volume Set, Chapman and Hall/CRC, 2016.
[21] Anna Maria Vegni, Loscri Valeria, A survey on vehicular social networks, Ieee Commun. Surv. Tutor. 17 (4) (2015) 2397–2419.
[22] Ahmad Faisal Abidin, Mario Kolberg, Amir Hussain, Improved Traffic Prediction Accuracy in Public Transport Using Trusted Information in Social Networks., Seventh York Doctoral Symposium on Computer Science & Electronics, 2014.
[23] Jian Cui, et al., Extraction of traffic information from social media interactions: methods and experiments. intelligent transportation systems (ITSC) IEEE 17th International Conference on, 2014, IEEE, 2014.
[24] Vito Latora, Marchiori Massimo, Is the Boston subway a small-world network? Phys. A Stat. Mech. Its Appl. 314 (1–4) (2002) 109–113.
[25] Katherine A. Seaton, Lisa M. Hackett, Stations, trains and small-world networks, Phys. A Stat. Mech. Its Appl. 339 (3–4) (2004) 635–644.
[26] Julian Sienkiewicz, Janusz A. Hołyst, Statistical analysis of 22 public transport networks in Poland, Phys. Rev. E 72 (4) (2005), 046127.
[27] Yong-Zhou Chen, Nan Li, He. Da-Ren, A study on some urban bus transport networks, Phys. A Stat. Mech. Its Appl. 376 (2007) 747–754.
[28] Jianhua Zhang, et al., Networked analysis of the Shanghai subway network, in China, Phys. A Stat. Mech. Its Appl. 390 (23–24) (2011) 4562–4570.
[29] Andor Háznagy, et al., Complex network analysis of public transportation networks: a comprehensive study International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2015, IEEE, 2015.
[30] Susan M. Grant-Muller, et al., Enhancing transport data collection through social media sources: methods, challenges and opportunities for textual data, Iet Intell. Transp. Syst. 9 (4) (2014) 407–417.
[31] Azizur Rahim, Xiangjie Kong, Feng Xia, Zhaolong Ning, Noor Ullah, Jinzhong Wang, Sajal K. Das, Vehicular social networks: a survey Pervasive and Mobile Computing, vol. 43, Elsevier, 2018, pp. 96–113.
[32] Dimitrios Tsiotas, Serafeim Polyzos, Introducing a new centrality measure from the transportation network analysis in Greece, Ann. Oper. Res. 227 (1) (2015) 93–117.
[33] Yew-Yih Cheng, et al., Measuring centralities for transportation networks beyond structures, in: Applications of Social Media and Social Network Analysis, Springer, Cham, 2015, pp. 23–39.
[34] Roger Guimera, et al., The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles, Proc. National Acad. Sci. U.S.A 102 (22) (2005) 7794–7799.
[35] Sybil Derrible, Christopher Kennedy, The complexity and robustness of metro networks, Phys. A Stat. Mech. Its Appl. 389 (17) (2010) 3678–3691.
[36] Grosenick, Scott. Real-time traffic prediction improvement through semantic mining of social networks, Diss (2012).
[37] Camille Roth, et al., A long-time limit for world subway networks, J. R. Soc. Interface (2012), rsif20120259.
[38] Panagiotis Angeloudis, David Fisk, Large subway systems as complex networks, Phys. A Stat. Mech. Its Appl. 367 (2006) 553–558.
[39] Kui Wang, Fu. Xiufen, Research on centrality of urban transport network nodes AIP Conference Proceedings., 1839, AIP Publishing, 2017, No. 1.
[40] Islam H. El-adaway, Ibrahim Abotaleb, Eric Vechan, Identifying the most critical transportation intersections using social network analysis, Transp. Plan. Technol. 41 (4) (2018) 353–374.
[41] Feifei Li, et al., On trip planning queries in spatial databases, in: International Symposium on Spatial and Temporal Databases, Springer, Berlin, Heidelberg, 2005.