algorithms

MDPI

Article

Identifying Hubs Through Influential Nodes in Transportation Network by Using a Gravity Centrality Approach

Worawit Tepsan [1], Aniwat Phaphuangwittayakul [1,2,*], Saronsad Sokantika [1] and Napat Harnpornchai [3]

[1] International College of Digital Innovation, Chiang Mai University, Chiang Mai 50200, Thailand; worawit.tepsan@cmu.ac.th (W.T.); saronsad.s@cmu.ac.th (S.S.)
[2] Zuse Institute Berlin, 14195 Berlin, Germany
[3] Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand; napateconcmu@gmail.com
* Correspondence: aniwat.ph@cmu.ac.th

Abstract: Hubs are strategic locations that function as central nodes within clusters of cities, playing a pivotal role in the distribution of goods, services, and connectivity. Identifying these vital hubs—through analyzing influential locations within transportation networks—is essential for effective urban planning, logistics optimization, and enhancing infrastructure resilience. This task becomes even more crucial in developing and less-developed countries, where such hubs can significantly accelerate urban growth and drive economic development. However, existing hub identification approaches face notable limitations. Traditional centrality measures often yield low variance in node scores, making it difficult to distinguish truly influential nodes. Moreover, these methods typically rely solely on either local metrics or global network structures, limiting their effectiveness. To address these challenges, we propose a novel method called Hybrid Community-based Gravity Centrality (HCGC), which integrates local influence measures, community detection, and gravity-based modeling to more effectively identify influential nodes in complex networks. Through extensive experiments, we demonstrate that HCGC consistently outperforms existing methods in terms of spreading ability across varying truncation radii. To further validate our approach, we introduce ThaiNet, a newly constructed real-world transportation network dataset. The results show that HCGC not only preserves the strengths of traditional local approaches but also captures broader structural patterns, making it a powerful and practical tool for real-world network analysis.

Keywords: transportation networks; connectivity; network analysis; hub identification; influential nodes; regional development; geospatial dataset

check for updates

Academic Editor: Eugene Levner

Received: 11 May 2025
Revised: 3 June 2025
Accepted: 7 June 2025
Published: 10 June 2025

Citation: Tepsan, W.; Phaphuangwittayakul, A.; Sokantika, S.; Harnpornchai, N. Identifying Hubs Through Influential Nodes in Transportation Network by Using a Gravity Centrality Approach. *Algorithms* **2025**, *18*, 356. https://doi.org/10.3390/a18060356

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Network analysis enriched with geospatial data offers a powerful framework for understanding spatial dynamics and supporting data-driven decision-making. The growing body of literature—reflected in works such as [1–3]—demonstrates the increasing availability of geospatial datasets and the advancement of analytical techniques, which have enabled widespread applications across diverse real-world domains. For instance, during Thailand's second wave of COVID-19, Peter Scully's GitHub project addressed the absence of official data by providing provincial adjacency information, thereby enabling more accurate modeling of disease transmission [4,5]. Adjacency-based networks have proven particularly useful in epidemiological studies. Ford et al. [6] developed the Spatiotemporal Epidemiological Modeler (STEM), an extensible framework for simulating disease spread across spatial and temporal scales. Similarly, Chinpong et al. [7] applied

spatiotemporal analysis to tuberculosis (TB) incidence in Thailand from 2011 to 2020, revealing significant spatial heterogeneity and temporal trends along international borders. Beyond epidemiology, Martínez Márquez and Patanè [8] demonstrated how combining spatial and mobility data—using methods such as eigenvector centrality and Deep Gravity modeling—can enhance insights into flow dynamics and network structures. In a different domain, Mansourihanis et al. [9] proposed a GIS-based framework for assessing land-use compatibility, showing improved accuracy and transparency over conventional approaches through a case study in Qaemshahr, Iran. Together, these studies underscore the versatility and impact of geospatial network analysis across various sectors, including public health, transportation, and urban planning.

These diverse applications of geospatial network analysis highlight its value in uncovering hidden patterns and guiding data-informed interventions across various domains. Among the many analytical goals within this framework, one particularly critical task is identifying key or influential nodes in a network. These nodes often serve as strategic connectors, control points, or high-impact agents whose presence significantly affects the structure, function, and resilience of the system. Whether in the context of disease transmission, urban mobility, or land-use planning, understanding which nodes exert the most influence can greatly enhance predictive modeling, resource allocation, and policy development. As a result, a wide range of algorithms has been developed to tackle this challenge, increasingly moving beyond simplistic topological measures. Researchers have introduced various techniques for influential node detection, as shown in [10,11].

In the context of urban studies, where nodes can represent locations, these influential nodes are often referred to as hubs. In general terms, a hub refers to a region or a location that plays a pivotal role within a network by serving as a central point where multiple connections converge. These hubs are critical to the efficiency and functionality of various systems, often acting as integrators of key infrastructure, enablers of high-throughput mobility, or reducers of overall transportation costs. Their strategic position within a network allows them to influence broader spatial patterns, facilitate the flow of goods and services, and support coordinated development. As such, understanding the role of hubs offers valuable insights into spatial organization, functional zoning, and regional growth dynamics. For instance, Wang et al. [12] investigated urban agglomerations along the Yangtze River and found that major hub cities like Wuhan exert substantial spillover effects on land-use intensity, underscoring their impact on interregional development and spatial planning.

Given the importance of hubs, numerous studies have proposed methods to identify them within spatial and mobility networks. Xu et al. [13] used a graph-based approach with POI density and a marginalized graph autoencoder to delineate urban functional zones. Kang et al. [14] developed a multilayer network model to detect stable hubs in Beijing's taxi network by capturing spatial-temporal correlations. Bellingeri et al. [15] analyzed the results of a real-world road network of Beijing after removing one or two nodes and accounting for the weighted structure of the network using classic binary node properties and network functioning measurements. Other works, such as Tian et al. [16] and Yuan et al. [17], focused on ridership-weighted models and supernetwork theory to reveal hierarchical and multimodal structures within transit systems. In addition, a multi-commodity, multimodal hierarchical hub location problem under demand uncertainty was explored by incorporating fuzzy demand, which contributed to optimizing both transportation costs and time [18]. Another approach proposed a hub location model based on a branch-and-cut algorithm and intermodal transportation, which sought to overcome several traditional assumptions in hub location modeling [19].

Hub identification plays a critical role in understanding and managing complex networks, yet existing methods still face notable challenges. Centrality-based methods have been extensively applied across disciplines—including physics, biology, telecommunications, computer science, sociology, infrastructure, and transportation—to detect influential nodes [20–23]. Common metrics such as degree, betweenness, and closeness centrality offer useful perspectives but often yield low-variance scores, limiting their ability to distinguish truly influential nodes. Moreover, these traditional metrics typically emphasize either local connections or global structure, frequently overlooking intermediate patterns and spatial relationships that are critical in real-world applications. As a result, they may fall short in capturing the complexity of regionally diverse transportation networks.

To overcome these limitations, a novel approach for hub identification in transportation networks called *Hybrid Community-based Gravity Centrality (HCGC)*, along with an additional local transportation network dataset, is presented. The key contributions of this study can be summarized as follows:

- We propose HCGC, a method for identifying influential nodes in various complex networks. HCGC not only calculates the importance scores of each node based on local connections but also considers the global structure of the network by integrating local centrality features, gravity-based spatial modeling, and community detection. Moreover, it provides high-variance scores that are straightforward to use for ranking influential nodes.
- The ThaiNet dataset, representing Thailand's transportation networks across 77 provinces and 928 districts, is introduced. It is constructed by contiguity-based adjacency, enabling the realistic modeling of regional connectivity.
- An application of HCGC to ThaiNet is demonstrated. HCGC is not only applicable to general datasets but is particularly well-suited for transportation networks.
- Extensive experiments and comprehensive analysis between the proposed approach and traditional centrality measures are conducted on nine public networks and the ThaiNet dataset.

2. Methods

When starting with many locations, identifying which ones should serve as hubs requires evaluating the importance of each location—often measured through centrality values. Centrality encapsulates a location's significance within a network, especially regarding its spatial positioning and interconnectivity.

In terms of mathematical representation, a network can be defined as $G = (V, E)$, where V is the set of nodes (or vertices), and E is the set of edges. In this context, each node represents a *location*, while edges represent *connections or relationships* between nodes—such as the existence of a direct road between two locations. The adjacency between nodes u and v is indicated by a_{uv}, where

$$a_{uv} = \begin{cases} 1, & \text{if there is an edge between node } u \text{ and node } v, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The total number of nodes in the network is denoted by N. The graph G can also be represented using the *adjacency matrix* A of dimension $N \times N$, where each row and column corresponds to a node in the network. The entry at the row corresponding to node u and the column corresponding to node v, denoted a_{uv}, indicates whether an edge exists between nodes u and v.

To assess the significance of each node within a network, a variety of quantitative measures have been developed. Some of these measures are classic centrality metrics,

which focus on characteristics such as connectivity, accessibility, or intermediary roles. Others, however, incorporate additional structural or contextual factors, like clustering behavior or influence diffusion. These methods are particularly valuable in contexts such as transportation, communication, or information networks, where understanding both local and global node influence is crucial for decisions regarding hub selection or network optimization. The following sections introduce several commonly used metrics for evaluating node importance, ranging from traditional centrality measures to more recent hybrid and enhanced models.

2.1. Degree Centrality (DC)

Degree centrality is one of the simplest and most intuitive centrality measures. It quantifies the importance of a node by counting the number of direct connections (or neighbors) it has. Nodes with high degree centrality are directly connected to many other nodes, indicating strong local influence. In the context of transportation [24], communication, or economic networks, degree centrality reflects the reachability between nodes. The degree centrality of a node v is given by Equation (2).

$$DC(v) = \deg(v) = \sum_{u \neq v} a_{uv} \tag{2}$$

where $DC(v)$ is the degree centrality of node v, and a_{uv} is the adjacency indicator, as defined in Equation (1).

2.2. Closeness Centrality (CC)

Closeness centrality measures how near a node is to all other nodes in the network. It is computed based on the inverse of the total shortest path distances from a given node to all other nodes. Nodes with high closeness centrality have shorter average path lengths to others, indicating they are more "central" in terms of their ability to efficiently reach, or be reached by, other nodes in the network. This metric is particularly useful for identifying nodes that are effective in distributing resources, information, or influence across the network. The closeness centrality of a node v is defined by Equation (3).

$$CC(v) = \frac{N-1}{\sum_{u \neq v} d(v,u)} \tag{3}$$

where $CC(v)$ is the closeness centrality of node v, N is the total number of nodes in the network, and $d(v,u)$ denotes the shortest path distance between nodes v and u. The summation $\sum_{u \neq v} d(v,u)$ is taken over all nodes u distinct from v.

2.3. Betweenness Centrality (BC)

Betweenness centrality measures a node's importance based on how often it appears on the shortest paths between pairs of other nodes in a network [25]. Nodes with high betweenness centrality serve as bridges, potentially controlling the flow of information, resources, or activity between different parts of the network. The formula for calculating betweenness centrality is given in Equation (4).

$$BC(v) = \sum_{u \neq v \neq w} \frac{\sigma_{uw}(v)}{\sigma_{uw}} \tag{4}$$

where $BC(v)$ is the betweenness centrality of node v, σ_{uw} is the total number of shortest paths from node u to node w, and $\sigma_{uw}(v)$ is the number of those paths that pass through node v.

2.4. Eigenvector Centrality (EC) from Adjacency Matrix

Eigenvector centrality measures a node's importance based on the importance of the nodes to which it is connected. Unlike simple degree centrality, which only counts direct connections, eigenvector centrality assigns higher scores to nodes that are connected to other highly central nodes. This measure is particularly useful for identifying influential nodes in the spread of economic influence or information within a network. Assume that A is the adjacency matrix of the network and λ is the largest eigenvalue of A. The calculation of eigenvector centrality is formally expressed in Equation (5).

$$EC(v) = \frac{1}{\lambda} \sum_{u \neq v} a_{uv} \, EC(u) \tag{5}$$

where $EC(v)$ is the eigenvector centrality of node v, and a_{uv} is the (v, u)-th entry of the adjacency matrix A.

2.5. Generalized Gravity Centrality (GGC)

GGC [26] is a gravity-based model that uses the degree of nodes as masses and includes local clustering coefficient to increase the spreading ability of the node. This approach shows better local information consideration of the node than conventional methods. The GGC is defined as follows:

$$\begin{aligned} GGC(v) &= \sum_{d_{uv} \leq R \& u} \frac{Sp(v) \times Sp(u)}{d_{uv}^2} \\ &= \sum_{d_{uv} \leq R \& u} \frac{e^{-\alpha C_v} \times k(v) \times e^{-\alpha C_u} \times k(u)}{d_{uv}^2} \end{aligned} \tag{6}$$

where R is a truncation radius for the gravity model [27]. $Sp(v)$ is spreading ability of node v. C_v is local clustering coefficient of node $v = 2n_v/(k(v)(k(v)-1))$. $k(v)$ is the degree of node v.

2.6. Hybrid Global Structure Model (HGSM)

HGSM [28] is an enhanced version of the Global Structure Model (GSM). HGSM integrates the Ks value into conventional self-influence (iSI) to increase the local impact of a node's DC. The self-influence (iSI) is then used to compute the global impact (iGI). The iGI factor calculates the sum of the neighbor ratios of the shortest path lengths by averaging the iSI values for directed and undirected nodes with Ks and DC values. The HGSM is defined as Equation (7).

$$\begin{aligned} HGSM(v) &= iSI(v) \times iGI(v) \\ &= e^{\frac{Ks(v) \times DC(v)}{N}} \times \sum_{v \neq u} \frac{e^{\frac{Ks(u) \times DC(u)}{N}}}{d_{uv}^{ceil(\log_2(ave_iSI))}} \end{aligned} \tag{7}$$

where $Ks(v)$ and $Ks(u)$ are K-shell decomposition of node v and node u. $DC(v)$ and $DC(u)$ are degree centrality of node v and node u, respectively. N is the total number of nodes.

2.7. Degree and Average Neighbor Degree (DC^+)

DC^+[29] is the improved method of the degree centrality to identify the node's influence. It measures the influence of nodes by a product of node degree and the average neighbor degree, which can be defined as Equation (8).

$$DC^+(v) = k(v) \times k_{nn,v} \tag{8}$$

where $k_{nn,v} = (\sum_{u \in N_v} k_u)/|N_u|$ is the average neighbor degree of the node v. N_v is the neighbor set of node v. $|N_v|$ is the size of node v.

3. Methodology

Inspired by HGSM [28] and the spreading ability of the node in GGC [26], we propose Hybrid Community-based Gravity Centrality (HCGC). Our method considers both the local and global information of each node in a network. Additionally, the method also includes the local modularity calculated by the Louvain algorithm due to its popularity for community detection [30]. This local modularity is used to improve the local influence scores (LI). HCGC is then computed by using LI as masses in the gravity model with a truncation radius R to capture the global information of the network. This approach has the potential to identify the vital hubs for transportation networks. The HCGC is defined by the following equation:

$$LI(v) = e^{\frac{LM(v)}{N}} \times DC^{+}(v) \times Ks(v) \tag{9}$$

where N is the number of nodes in a network. $LM(v)$ is the local modularity calculated by the community detection algorithm, which measures how well each individual node fits within its assigned community. Equation (10) defines the local modularity.

$$LM(v) = \frac{E_{internal}(v)}{E_{total}(v)} = \frac{\sum_{u \in N(v)} \delta(c_v, c_u)}{k_v} \tag{10}$$

where $E_{internal}(v)$ counts edges connecting node n to nodes in the same community. $E_{total}(v)$ is the degree of node v. $N(v)$ is the set of neighbors of node v. $\delta(c_v, c_u)$ is 1 if nodes v and u are in the same community, 0 otherwise. k_v is the degree of node n.

The community assignment $\delta(c_v, c_u)$ is calculated by the Louvain algorithm which is widely used for community detection in networks [30]. The Louvain algorithm maximizes global modularity Q to determine communities. This global modularity Q measures the quality of the entire community partition, while local modularity measures how well each individual node fits within its assigned community. The global modularity Q can be expressed as:

$$\begin{aligned} Q &= \frac{1}{2m} \sum_{v,u} \left[A_{uv} - \frac{k_v k_u}{2m} \right] \delta(c_v, c_u) \\ &= \frac{1}{2m} \left[\sum_{v,u} - \frac{\sum_v k_v \sum_u k_u}{2m} \right] \delta(c_v, c_u) \end{aligned} \tag{11}$$

$$\delta(c_v, c_u) = \begin{cases} 0, \ c_v, c_u \text{ belong to different community} \\ 1, \ c_v, c_u \text{ belong to same community} \end{cases}$$

where A_{uv} is the weight of the edge between nodes v and u, whose value is 1 for unweighted networks. k_v and k_u are the sums of weights of edges attached to nodes v and u. m is the sum of all edge weights in the graph. c_v is the community to which node v belongs.

HCGC represents the impact of neighbor nodes within a truncation radius R following gravity-based centrality measures [26,29].

$$HCGC(v) = \sum_{d_{uv} \leq R, v \neq u} \frac{LI(v) \times LI(u)}{d_{uv}^2} \tag{12}$$

Algorithm 1 shows the pseudocode of the HCGC.

Algorithm 1 The algorithm of HCGC measure

Input: Network $G(V, E)$ with nodes labels $0, 1, \ldots, N$, R
Output: $HCGC$ values for all nodes

Calculate LM using method in Equation (10) for all nodes
Calculate degree and average neighbor degree DC^+ for all nodes
Calculate Ks using K-shell decomposition method for all nodes
for each node v in V **do**
    $LI(v) \leftarrow e^{\frac{LM(v)}{N}} \times DC^+(v) \times Ks(v)$ // Calculate LI for node v according Equatio (9)
end for
for each node v in V **do**
    $HCGC \leftarrow 0$
    $R_v \leftarrow$ a set of R-neighborhood nodes of node v
    for each node u in R_v **do**
        $HCGC+ = \frac{LI(v) \times LI(u)}{d_{uv}^2}$ // Calculate the $HCGC$ according to Equation (12)
    end for
end for
return $HCGC$

3.1. SIR Spreading Dynamics

The SIR (Susceptible, Infected, Recovered) spreading dynamics model is a widely used methodology for assessing the node importance within a network based on their spreading ability. The predicted ranking of nodes based on their spreading ability can be obtained using this model. With the SIR model, each node in a network is categorized into one of three distinct states: *Susceptible* (S), $Infected$ (I), or *Recovered* (R). *Susceptible* (S) nodes represent individuals who have not yet contracted the infection but remain vulnerable to transmission. $Infected$ (I) nodes constitute those individuals who have acquired the infection and possess the capacity to transmit the disease to susceptible individuals within their network. *Recovered* (R) nodes denote individuals who were formerly infected but have subsequently recovered from the disease and are no longer capable of transmission.

The fundamental principle underlying this method involves these steps: (1) One node is selected as an infected node, while all other nodes are initially classified as susceptible nodes. (2) At each time step, every infected node has the probability β of transmitting the infection to its neighbor nodes. (3) The infected nodes transition to the recovered state with probability λ (which is set to 1 in this study). Different infection probabilities (β) are applied for this experiment. We set them as $\beta \in [0.5\beta_c, 1.5\beta_c]$. β_c is the spreading threshold of the SIR model for each network, defined as the following equation.

$$\beta_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \tag{13}$$

where $\langle k \rangle$ denotes the average degree of the network, $\langle k^2 \rangle$ represents the second-order average degree of the network. This process terminates when the network no longer contains any infected nodes.

Based on this model, the influence of a given node i can be estimated as $F(i) = N_r/N$, where N_r represents the total number of recovered nodes in the network upon reaching the steady state of the SIR spreading dynamics process. To achieve a more accurate approximation of the true ranking of node spreading ability, the final spreading influence for each initially infected node is determined by calculating the statistical average over 1000 iterations, $N = 1000$. The ranking results of network nodes' SIR spreading abilities are obtained through the implementation of the Epidemic on Networks (EoN) module [31].

3.2. Kendall's Tau (τ)

The accuracy of different algorithms is subsequently evaluated through the application of Kendall's tau correlation coefficient [32]. Kendall's tau represents a widely adopted statistical measure for quantifying the strength of the correlation between two sequences. When considering two sequences X and Y with the same length, two pairs of elements (x_i, y_i) and (x_j, y_j) are selected from corresponding positions within these sequences. These pairs are classified as concordant if $x_i > x_j$ and $y_i > y_j$, or if $x_i < x_j$ and $y_i < y_j$. Conversely, the pairs are considered discordant if $x_i > x_j$ and $y_i < y_j$, or if $x_i < x_j$ and $y_i > y_j$. The pairs are neither concordant nor discordant if $x_i = x_j$ or $y_i = y_j$. Denoting the number of concordant and discordant pairs as N_c and N_d, respectively, Kendall's tau between sequences X and Y can be expressed as Equation (14).

$$\tau(X, Y) = \frac{N_c - N_d}{N(N-1)/2} \tag{14}$$

where $\tau \in [-1, 1]$. A value of $\tau = 1$ indicates a perfect match between the rankings of sequences X and Y, whereas $\tau = -1$ signifies a complete inversion of rankings. When $\tau = 0$, the rankings of the two sequences X and Y are completely independent. N_c and N_d denote the number of concordant and discordant pairs, respectively. N is the number of total nodes in the network.

For evaluating the accuracy of various measures, the spreading capability of nodes is initially determined through the implementation of the SIR model. Subsequently, a ranking sequence is generated by ordering node influence as determined through algorithmic simulation. The final step involves the computation of Kendall's tau (τ) between these two sequences. Measures with τ values close to 1 demonstrate the highest performance in accurately identifying influential nodes within the network.

A critical advantage of SIR models in transportation network analysis is their natural compatibility with network topology considerations. The adjacency matrix of transportation networks, such as road networks, directly influences the diffusion dynamics, as the spread of any phenomenon is constrained by the physical connectivity of road segments. This topological awareness allows SIR models to capture realistic diffusion patterns that respect the geometric constraints of road infrastructure.

3.3. Imprecision Function

The imprecision function is another evaluation method to assess the spreading ability of influential nodes. This function compares a fraction of the nodes (p) which differs from Kendall's tau that compares the rank of all nodes with the ranking obtained by the SIR model [33]. The imprecision function is denoted as Equation (15).

$$\epsilon(p) = \frac{F_m(\vartheta)}{F_{eff}(\vartheta)} \tag{15}$$

where $F_m(\vartheta)$ represents the average spreading ability when spreading originates from nodes ranked highest by the given measure m. $F_{eff}(\vartheta)$ represents the average spreading ability of the actual top ϑ nodes in the SIR model. To compare the spreading ability in the early stage, we run the SIR model for 20 iterations to obtain the average scores.

4. Datasets

Two main datasets are used in this study to evaluate the performance of the proposed method. The first is a publicly available real-world network dataset, while the second is a custom-generated network dataset developed for this research and referred to as ThaiNet.

4.1. Public Real Networks

A total of nine public real-world networks are used in this study, including two transportation networks (USAir and Euroroad), three social networks (Dolphins, Blogs, and Karate), three communication networks (Email, EEC, and GDciting), and one biological network (Celegan), as described in Table 1.

Table 1. The name, type, and description for each network in public real networks dataset.

Network	Type	Description
Dolphins	Social	Bottlenose dolphin social interactions in New Zealand waters
USAir	Transportation	Undirected transportation network representing the US Air route system from 1997
EEC	Communication	European researchers exchange of emails
Email	Communication	Email communication network where users send emails and communicate with each other
Euroroad	Transportation	European road network system spanning multiple countries
Blogs	Social	A communication network of blogs
Karate	Social	Social network of friendships between 34 members of a karate club at a US university in the 1970s.
GDciting	Communication	A network of meeting articles that appeared in 1994–2000
Celegan	Biological	Caenorhabditis elegans molecular interaction networks

These datasets are publicly accessible through NETWORKREPOSITORY (https://networkrepository.com/networks.php, accessed on 14 February 2025) and KONECT (http://konect.cc/networks/, accessed on 14 February 2025). The basic properties of each network are summarized in Table 2.

Table 2. Basic properties of public real networks, including number of nodes (N), edges (E), average degree ($\langle k \rangle$), shortest path length ($\langle d \rangle$), clustering coefficient ($\langle c \rangle$), and epidemic threshold (β_c).

Network	N	E	$\langle k \rangle$	$\langle d \rangle$	$\langle c \rangle$	β_c
Dolphins	62	159	5.129	3.357	0.259	0.172
USAir	332	2126	12.807	2.738	0.625	0.023
EEC	986	16,064	32.584	2.587	0.407	0.014
Email	1133	5451	9.622	3.606	0.220	0.057
Euroroad	1174	1417	2.414	2.418	0.017	0.500
Blogs	1222	16,714	27.355	2.738	0.320	0.012
Karate	34	78	4.588	2.408	0.571	0.148
GDciting	259	640	4.942	1.525	0.230	0.144
Celegan	248	468	3.774	1.580	0.071	0.208

4.2. ThaiNet Dataset

4.2.1. Steps for Generating Sample Transportation Network

The steps include collecting the central location of each region, constructing adjacency relationships based on regional boundaries, and generating an adjacency matrix for network analysis, as illustrated in Figure 1.



Figure 1. Process of generating regional datasets, which includes collecting location data, constructing adjacency pairs based on regional boundaries, and finally generating adjacency matrix.

Region Locations Data Collection

Data collection begins with identifying the central location for each region. Then, latitude and longitude coordinates of these central points are collected and compiled into a CSV file, which includes the names and coordinates of each location.

Adjacency Region Pair Construction

The adjacency of regions is determined through a series of algorithms. Initially, boundary overlap is checked to filter out non-adjacent district pairs quickly (see Algorithm 2). Subsequently, detailed proximity checks are performed to confirm adjacency (see Algorithm 3). These validated adjacencies are then used to construct pairs for further analysis (see Algorithm 4).

Algorithm 2 Check Extremities Boundary Boxes Overlap

procedure CHECKBOXOVERLAP($bound1, bound2, \text{threshold} = 0.01$)
  $min_lon1, min_lat1, max_lon1, max_lat1 \leftarrow bound1$
  $min_lon2, min_lat2, max_lon2, max_lat2 \leftarrow bound2$
  $lat_overlap \leftarrow (max_lat1 \geq min_lat2 - \text{threshold}) \wedge (max_lat2 \geq min_lat1 - \text{threshold})$
  $lon_overlap \leftarrow (max_lon1 \geq min_lon2 - \text{threshold}) \wedge (max_lon2 \geq min_lon1 - \text{threshold})$
  return $lat_overlap \wedge lon_overlap$
end procedure

Algorithm 3 Check if Two Regions Are Adjacent Based on Overlapping Boundary Proximity

Require: $region1, region2$, distance_threshold $= 0.05$
$box_overlap \leftarrow$ CHECKBOXOVERLAP($bound1, bound2$)
if $box_overlap$ **then**
  Perform detailed proximity checks to confirm adjacency
  return True if adjacent, False otherwise
end if
return False

Algorithm 4 Construct Adjacency Region Pairs

Initialize empty list AdjacencyPair = [].
for each region1 in allregions **do**
  for each region2 in allregions **do**
    $isAdjacent \leftarrow$ CHECKADJACENCY($region1, region2$)
    if $isAdjacent$ **then**
      AdjacencyPair.append({region1, region2})
    end if
  end for
end for

Adjacency Generation

Following the establishment of adjacency pairs, we generate matrices to analyze spatial relationships. The adjacency matrix records regions that are directly adjacent. Specifically, we measure the distance between each pair of adjacent regions back and forth, and use the average of these measurements as the distance entry in our adjacency matrix (see Algorithms 5).

Algorithm 5 Generate Adjacency Matrix

```
Initialize matrix A as an N × N array filled with zeros
for each region1, region2 in AdjacencyPair do
    A[region1_index, region2_index] ← 1
end for
```

4.2.2. Overview of the ThaiNet Dataset

The steps for generating a sample transportation network are applied to Thailand's regions, which include 77 provinces and 928 districts. As a result, ThaiNet consists of two main networks—one for the provinces and another for the districts. Each dataset will include:

- A list of regions along with their central locations,
- Pairs of adjacent regions, and
- An adjacency matrix representing the boundary connections between regions.

Province Network in ThaiNet

Thailand comprises 77 provinces, for which data were collected using a combination of the described methodology and manual efforts. The collected data include:

- A compiled list of all 77 provinces of Thailand, along with the precise locations of their respective Muang District Offices. These offices serve as the administrative hubs for their provinces. Notably, for Bangkok, the Democracy Monument serves as the representative location, as detailed in Table 3.
- An inventory of a total of 358 paired adjacent provinces, comprising 179 pairs of non-switched order pairs, which provides insights into the distances between them. This dataset highlights the geographical proximity and connectivity among the provinces. Detailed information on adjacent provinces is provided in Table 4.
- An Adjacency Matrix is tailored to reflect the spatial relationships among the 77 provinces, as shown in Table 5.

Table 3. List of 77 provinces with corresponding coordinates of their central location.

Order	Province	Latitude	Longitude
0	Bangkok	13.765171	100.539168
1	Samut Prakan	13.600400	100.597082
⋮	⋮	⋮	⋮
76	Narathiwat	6.425175	101.823515

Table 4. List of 357 adjacent province pairs.

Order	**Origin**	**Destination**
0	Bangkok	Samut Prakan
1	Bangkok	Nonthaburi
⋮	⋮	⋮
357	Narathiwat	Yala

Table 5. Adjacency matrix illustrating the connectivity between provinces such as Bangkok, Samut Prakan, and Narathiwat.

	Bangkok	**Samut Prakan**	···	**Narathiwat**
Bangkok	0	1	···	0
Samut Prakan	1	0	···	0
⋮	⋮	⋮	⋮	⋮
Narathiwat	0	0	···	0

District Network ThaiNet

Similarly to Provinces data, we generate dataset for 928 Thailand districts. The following data have been collected for each district:

- A list of all districts, along with the locations of their District Offices. Each district is referenced by a tuple consisting of the district name and its respective province (i.e., (district, province)). The location used to represent each district is the main office within that district. These data will be collected in the format presented in Table 6.
- A list of 6192 pairs of adjacent districts (3096 pairs, not counting reverse originals and destinations) will be generated in the format of Table 7.
- An Adjacency Matrix captures the spatial dimensions and relationships among the 928 districts, as presented in Table 8.

Table 6. List of districts in Thailand along with the geographic coordinates (latitude and longitude) of their central locations.

Order	**District Tuple**	**Latitude**	**Longitude**
0	(Phra Nakhon, Bangkok)	13.765096	100.499357
1	(Dusit, Bangkok)	13.777345	100.520936
⋮	⋮	⋮	⋮
927	(Cho-airong, Narathiwat)	6.225381	101.811457

Table 7. Lists pairs of adjacent districts, detailing each origin-destination relationship to showcase the direct geographic or administrative connectivity between districts.

Order	**Origin**	**Destination**
0	(Mae Sariang, Mae Hong Son)	(Hot, Chiang Mai)
1	(Mae Sariang, Mae Hong Son)	(Omkoi, Chiang Mai)
⋮	⋮	⋮
6191	(Wang Sombun, Sa Kaeo)	(Soi Dao, Chanthaburi)

Table 8. Adjacency matrix for districts in Thailand, where a 1 indicates direct adjacency between districts and a 0 denotes no adjacency.

	(Phra Nakhon, Bangkok)	**(Dusit, Bangkok)**	⋯	**(Cho-airong, Narathiwat)**
(Phra Nakhon, Bangkok)	0	1	⋯	0
(Dusit, Bangkok)	1	0	⋯	0
⋮	⋮	⋮	⋮	⋮
(Cho-airong, Narathiwat)	0	0	⋯	0

The basic network properties of the province and district networks in the ThaiNet dataset are summarized in Table 9.

Table 9. Basic properties of the province and district networks in the ThaiNet dataset, including number of nodes (N), edges (E), average degree ($\langle k \rangle$), shortest path length ($\langle d \rangle$), clustering coefficient ($\langle c \rangle$), and epidemic threshold (β_c).

Network	N	E	$\langle k \rangle$	$\langle d \rangle$	$\langle c \rangle$	β_c
Province	77	179	4.649	5.586	0.505	0.230
District	928	3057	6.588	16.594	0.491	0.149

5. Results and Discussion

This section demonstrates the performance of HCGC and baseline measures through SIR spreading dynamics experiments on both public real networks and ThaiNet datasets. These baseline measures consist of degree centrality (DC), closeness centrality (CC), betweenness centrality (BC), eigenvector centrality (EC), DC^{+} [29], GGC [26], LGI [34], $DCGM^{++}$ [29], HGSM [28], NPIC [35], and NHM [33]. The results obtained from SIR can be used to identify the influential nodes in the complex network. All experiments were run on an Intel(R) Core(TM) i5-12450H CPU 2.00 GHz processor with 16 GB of memory.

5.1. SIR Spreading Dynamic Results

Table 10 demonstrates the efficiency of our method compared with baseline measures through a truncation radius $R = 2$ and infection probability $\beta = \beta_c$. Although $DCGM^{++}$, which is another gravity centrality measure, demonstrates the highest accuracy in Dolphins and USAir, the results show that HCGC outperforms the existing approaches on six networks including EEC, Email, Euroroad, Blogs, GDciting, and Celegan. Moreover, our measure achieves the second-best performance on Dolphins and USAir networks. Notably, HCGC shows good performance on transportation networks (USAir and Euroroad). Additionally, the results show that DC^{+} demonstrates better efficiency than DC on every network, which is why we chose DC^{+} instead of traditional DC as a component in our method.

Figure 2 presents the spreading capability of different measures on nine networks via Kendall's correlation coefficient. The figure consists of the accuracy of various infection probabilities (β) ranging from $0.5\beta_c$ to $1.5\beta_c$ and $R = 2$ for each network. HCGC shows better performance in Dolphins, USAir, EEC, Email, Euroroad, Blogs, GDciting, and Celegan when the infection probability is greater than β_c. Moreover, the accuracy of our measure consistently increases compared to others. In particular, in the USAir, EEC, Email, Blogs, GDciting, and Celegan networks, the accuracy curve shows an upward trend compared to other measures that remain relatively stable.

Table 10. The Kendall's correlation coefficient of different measures over public real networks. Where $\beta = \beta_c$ and $R = 2$ for a truncation radius. The measure with highest accuracy for each network is highlighted in bold and the measure with the second best is underlined.

Network	DC	CC	BC	EC	DC^{+}	GGC	LGI	DCGM^{++}	HSGM	NPIC	NHM	HCGC
Dolphins	0.8274	0.6260	0.5590	0.6587	0.8765	0.7963	0.7984	**0.9148**	0.8566	0.8661	0.8619	0.9074
USAir	0.7679	0.8008	0.5758	0.8803	0.8900	0.7471	0.8517	**0.8966**	0.7886	0.8068	0.8721	0.8922
EEC	0.8524	0.8504	0.7254	0.9042	0.9052	0.8404	0.8577	0.9007	0.7648	0.8699	0.8969	**0.9058**
Email	0.8000	0.8079	0.6421	0.8634	0.8937	0.8077	0.8517	0.8873	0.8653	0.8326	0.8955	**0.9067**
Euroroad	0.6217	0.5933	0.4173	0.5901	0.8217	0.7799	0.7729	0.8193	0.6967	0.7690	0.7944	**0.8662**
Blogs	0.8603	0.7814	0.6797	0.8584	0.8889	0.8123	0.8454	0.8979	0.7169	0.8759	0.8893	**0.9018**
Karate	0.7890	0.7505	0.6573	0.8441	0.8146	0.7226	0.7178	**0.9198**	0.8667	0.8477	0.9064	0.8825
GDciting	0.7686	0.7777	0.5505	0.8263	0.9099	0.7727	0.8453	0.9051	0.8636	0.8410	0.9036	**0.9241**
Celegan	0.7246	0.8182	0.5462	0.8069	0.8664	0.8022	0.7958	0.867	0.8687	0.8225	0.8870	**0.9118**

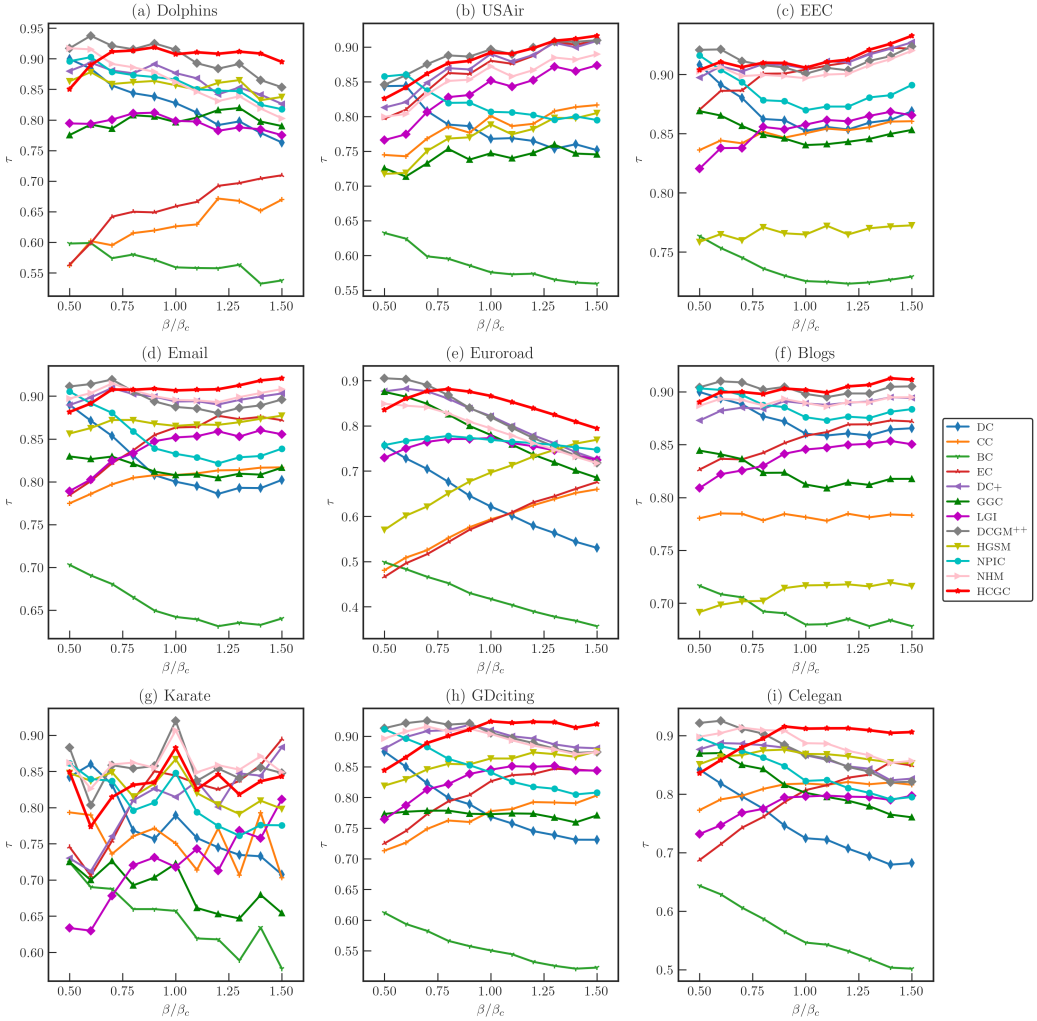


Figure 2. Comparison of Kendall's correlation coefficient of different measures on public real networks dataset in range of $\beta = [0.5\beta_c, 1.5\beta_c]$ and $R = 2$.

Table 11 presents the efficiency of our method compared with baseline measures through a truncation radius $R = 3$ and infection probability $\beta = \beta_c$. Similar to the previous setting with a truncation radius $R = 2$, the HCGC outperforms existing works in most networks. However, DC^{+} shows better performance in the EEC network. Comparing the results between $R = 2$ and $R = 3$, $R = 3$ shows better accuracy on USAir, Euroroad, Karate, and Celegan networks. This demonstrates that HCGC performs effectively on transportation networks.

Table 11. The Kendall's correlation coefficient of different measures over public real networks. Where $\beta = \beta_c$ and $R = 3$ for a truncation radius. The measure with highest accuracy for each network is highlighted in bold and the measure with the second best is underlined.

Network	DC	CC	BC	EC	DC^+	GGC	LGI	$DCGM^{++}$	HSGM	NPIC	NHM	HCGC
Dolphins	0.8274	0.6260	0.5590	0.6587	0.8765	0.7709	0.7984	**0.9148**	0.8566	0.8661	0.8619	0.9000
USAir	0.7679	0.8008	0.5758	0.8803	0.8900	0.7077	0.8517	**0.8964**	0.7886	0.8068	0.8866	0.8956
EEC	0.8524	0.8504	0.7254	0.9042	**0.9052**	0.8226	0.8577	0.8985	0.7648	0.8699	0.8969	0.9038
Email	0.8000	0.8079	0.6421	0.8634	0.8937	0.7770	0.8517	0.8804	0.8653	0.8326	0.8898	**0.8989**
Euroroad	0.6217	0.5933	0.4173	0.5901	0.8217	0.8167	0.7729	0.8389	0.6967	0.7690	0.8060	**0.8833**
Blogs	0.8603	0.7814	0.6797	0.8584	0.8889	0.7911	0.8454	0.8970	0.7169	0.8759	0.8949	**0.9012**
Karate	0.7890	0.7505	0.6573	0.8441	0.8146	0.7034	0.7178	**0.9126**	0.8667	0.8477	0.8955	0.8982
GDciting	0.7686	0.7777	0.5505	0.8263	0.9099	0.7593	0.8453	0.9089	0.8636	0.8410	0.9095	**0.9226**
Celegan	0.7246	0.8182	0.5462	0.8069	0.8664	0.8095	0.7958	0.8776	0.8687	0.8225	0.9038	**0.9137**

Figure 3 illustrates the spreading capability of different measures through the Kendall's correlation coefficient with a truncation radius $R = 3$ and infection probabilities $\beta \in [0.5\beta_c, 1.5\beta_c]$. Similar to the results of $R = 2$, HCGC outperforms the existing approaches on most of the networks when $\beta > \beta_c$, except for the Karate network.

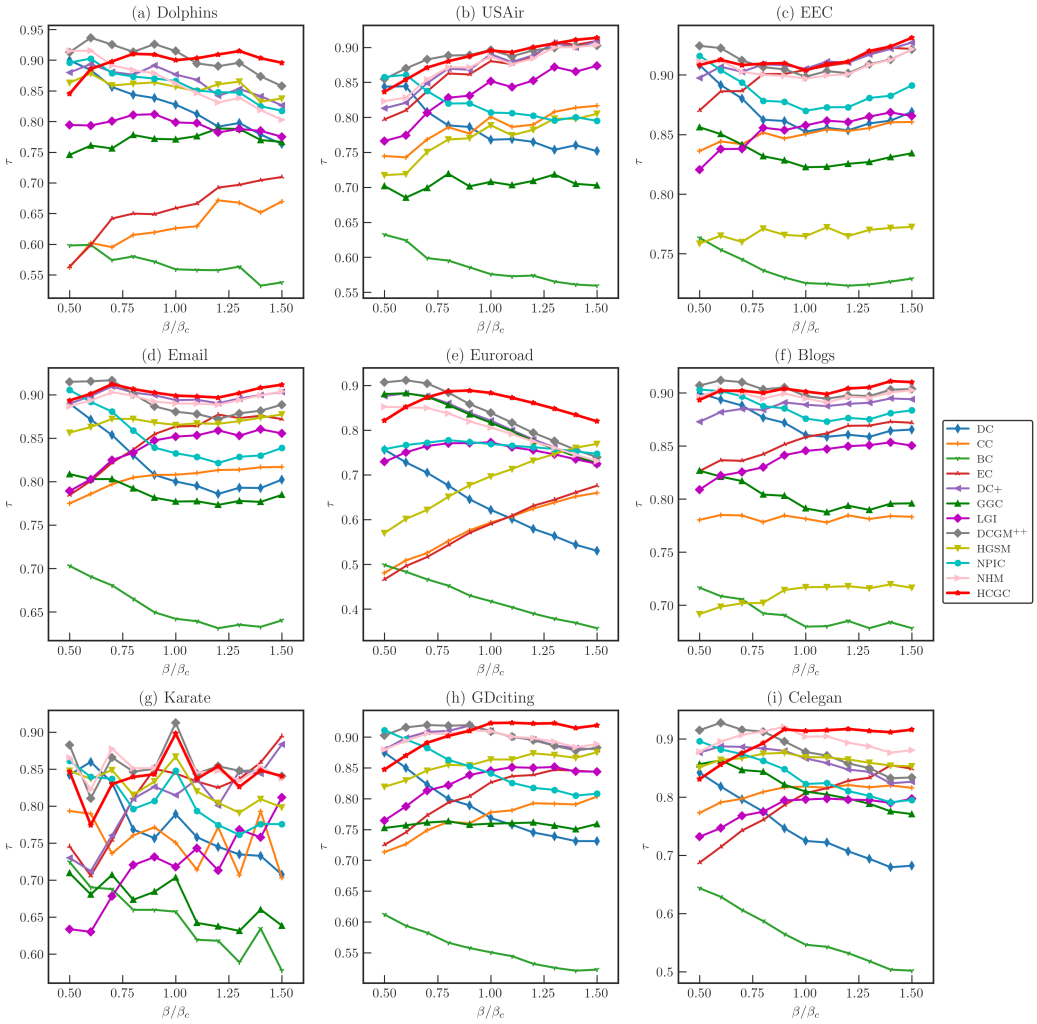


Figure 3. Comparison of Kendall's correlation coefficient of different measures on the public real networks dataset in range of $\beta = [0.5\beta_c, 1.5\beta_c]$ and $R = 3$.

5.2. Imprecision Function Results

Table 12 demonstrates the imprecision function values of baseline measures and our proposed algorithm on the public real networks dataset. HCGC with $R = 2$ achieves the highest values compared to others on four networks including Email, Euroroad, Blogs, and Celegan. The proposed measure slightly underperforms the EC on the EEC network.

Table 12. Imprecision function of different measures over public real networks, where $R = 2$ for a truncation radius. The measure with highest accuracy for each network is highlighted in bold and the measure with the second best is underlined.

Network	DC	CC	BC	EC	DC^{+}	GGC	LGI	$DCGM^{++}$	HSGM	NPIC	NHM	HCGC
Dolphins	0.1031	0.2322	0.3082	0.0932	0.1022	0.1326	0.1829	0.0909	0.0863	0.0908	**0.0810**	0.1007
USAir	0.1250	**0.1226**	0.2358	0.1351	0.1293	0.1385	0.1389	0.1247	0.1286	0.1262	0.1275	0.1340
EEC	0.1944	0.1829	0.2567	**0.1681**	0.1714	0.1931	0.1760	0.1764	0.2132	0.1815	0.1779	0.1683
Email	0.2254	0.2154	0.2734	0.2208	0.2037	0.2147	0.2240	0.2038	0.2094	0.2151	0.2034	**0.2030**
Euroroad	0.1787	0.3034	0.3750	0.2104	0.1208	0.1535	0.2176	0.1272	0.1685	0.1430	0.1437	**0.1127**
Blogs	0.2093	0.2606	0.2987	0.2044	0.1836	0.2243	0.2002	0.1922	0.2132	0.1970	0.1937	**0.1823**
Karate	0.0812	0.1317	0.0708	0.0883	0.0783	0.0876	0.3177	0.0646	0.0812	**0.0635**	0.0635	0.0783
GDciting	0.1911	0.1897	0.3679	0.1512	0.1458	0.1812	**0.1455**	0.1600	0.1656	0.1626	0.1476	0.1484
Celegan	0.2128	0.1915	0.3551	0.1058	0.1095	0.1471	0.1433	0.1206	0.1487	0.1374	0.1193	**0.0991**

Table 13 represents the imprecision function values of baseline measures and our proposed algorithm on public real networks dataset. HCGC with $R = 3$ achieves the highest values compared to others on four networks including Email, Euroroad, Blogs and Celegan. Comparing between the proposed method with $R = 2$ and $R = 3$, HCGC with $R = 3$ shows better evaluation on USAir, Euroraod, Karate, and GDciting networks.

Table 13. Imprecision function of different measures over public real networks, where $R = 3$ for a truncation radius. The measure with highest accuracy for each network is highlighted in bold and the measure with the second best is underlined.

Network	DC	CC	BC	EC	DC^{+}	GGC	LGI	$DCGM^{++}$	HSGM	NPIC	NHM	HCGC
Dolphins	0.1031	0.2322	0.3082	0.0932	0.1022	0.1381	0.1829	0.0863	0.0863	0.0908	**0.0810**	0.0977
USAir	0.1250	**0.1226**	0.2358	0.1351	0.1293	0.1372	0.1389	0.1247	0.1286	0.1262	0.1282	0.1340
EEC	0.1944	0.1829	0.2567	**0.1681**	0.1714	0.1932	0.1760	0.1764	0.2132	0.1815	0.1770	0.1683
Email	0.2254	0.2154	0.2734	0.2208	**0.2037**	0.2177	0.2240	0.2072	0.2094	0.2151	0.2061	0.2058
Euroroad	0.1787	0.3034	0.3750	0.2104	0.1208	0.1444	0.2176	0.1238	0.1685	0.1430	0.1399	**0.1091**
Blogs	0.2093	0.2606	0.2987	0.2044	0.1836	0.2249	0.2002	0.1939	0.2132	0.1970	0.1917	**0.1831**
Karate	0.0812	0.1317	0.0708	0.0883	0.0783	0.0876	0.3177	0.0646	0.0812	**0.0635**	0.0788	0.0766
GDciting	0.1911	0.1897	0.3679	0.1512	0.1458	0.1903	**0.1455**	0.1588	0.1656	0.1626	0.1492	0.1474
Celegan	0.2128	0.1915	0.3551	0.1058	0.1095	0.1660	0.1433	0.1242	0.1487	0.1374	0.1151	**0.1014**

Figures 4 and 5 illustrate the imprecision function values across different measures. The horizontal axis refers to the fractions of top nodes in the network (p) ranging from 0.01 to 0.2. The vertical axis shows the calculated values of the imprecision function based on the specific algorithm $\epsilon(p)$. Based on the curves, they show that our method presents significantly high performance (lower is better) compared to the traditional methods on the Euroroad network. Moreover, HCGC has curves that are similar to $DCGM^{++}$ and NHM on USAir, Email, Blogs, Karate, and GDCiting networks.

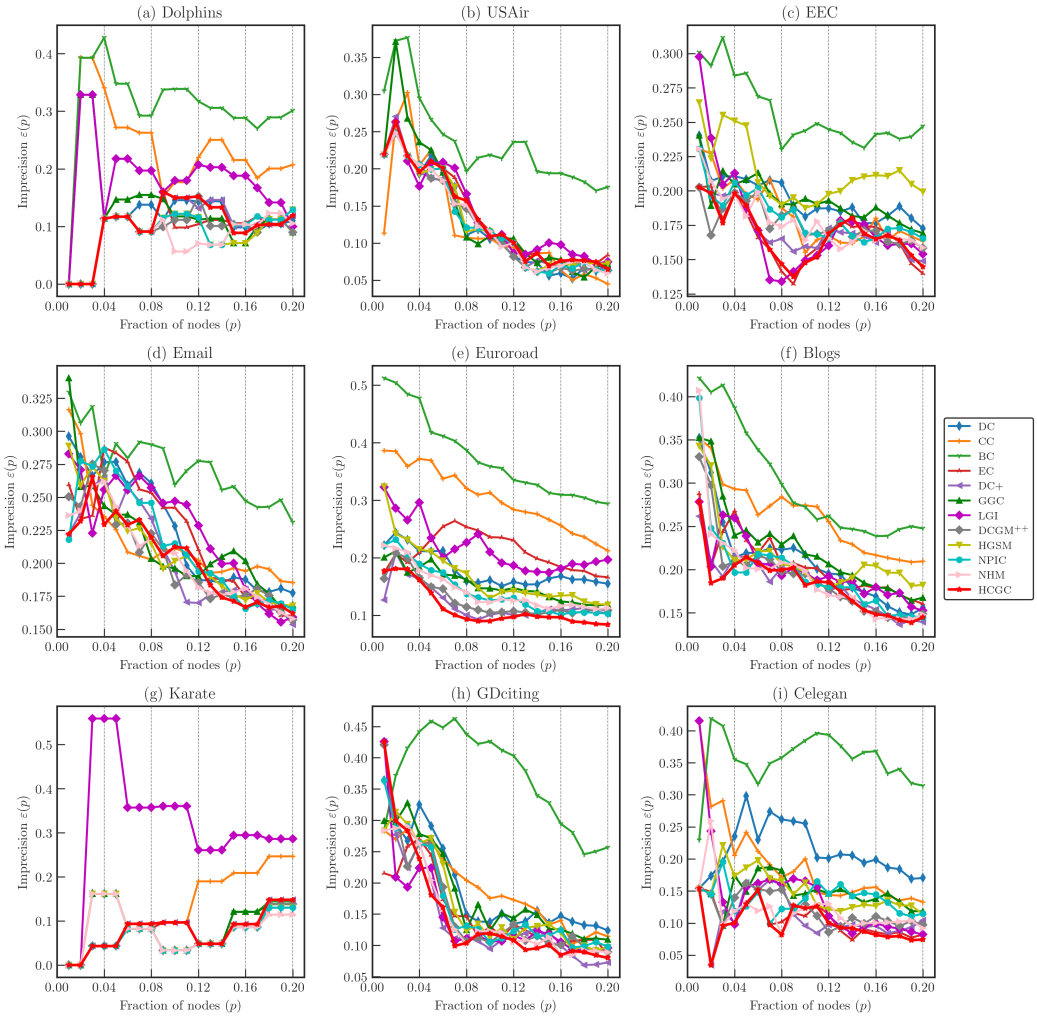


Figure 4. Comparison of imprecision function of different measures on public real networks dataset with $\beta = \beta_c$ and $R = 2$.

Table 14 presents the average imprecision function values across all networks for different measures. The results show that our proposed method outperforms other methods with the lowest imprecision function values for both truncation radii $R = 2$ and $R = 3$. DC^+ is the second-best performing measure.

Table 14. Imprecision function average of different measures over public real networks, where $R = 2$ and $R = 3$ for a truncation radius. The measure with highest accuracy for each network is highlighted in bold and the measure with the second best is underlined.

Measure	DC	CC	BC	EC	DC^+	GGC	LGI	$DCGM^{++}$	HSGM	NPIC	NHM	HCGC
R = 2	0.1690	0.2033	0.2824	0.1530	<u>0.1383</u>	0.1636	0.1940	0.1400	0.1572	0.1463	0.1414	**0.1363**
R = 3	0.1690	0.2033	0.2824	0.1530	<u>0.1383</u>	0.1666	0.1940	0.1400	0.1572	0.1463	0.1408	**0.1359**

5.3. Centrality Measurement in the ThaiNet Dataset

Since HCGC is an approach that is applicable to unweighted networks, the weights, such as distance, between each node were disregarded during construction of the ThaiNet dataset. Moreover, using a binary network preserves the simplification that effectively captures the fundamental structural properties of networks, such as network topology, and provides computational advantages that enable large-scale urban network analysis.

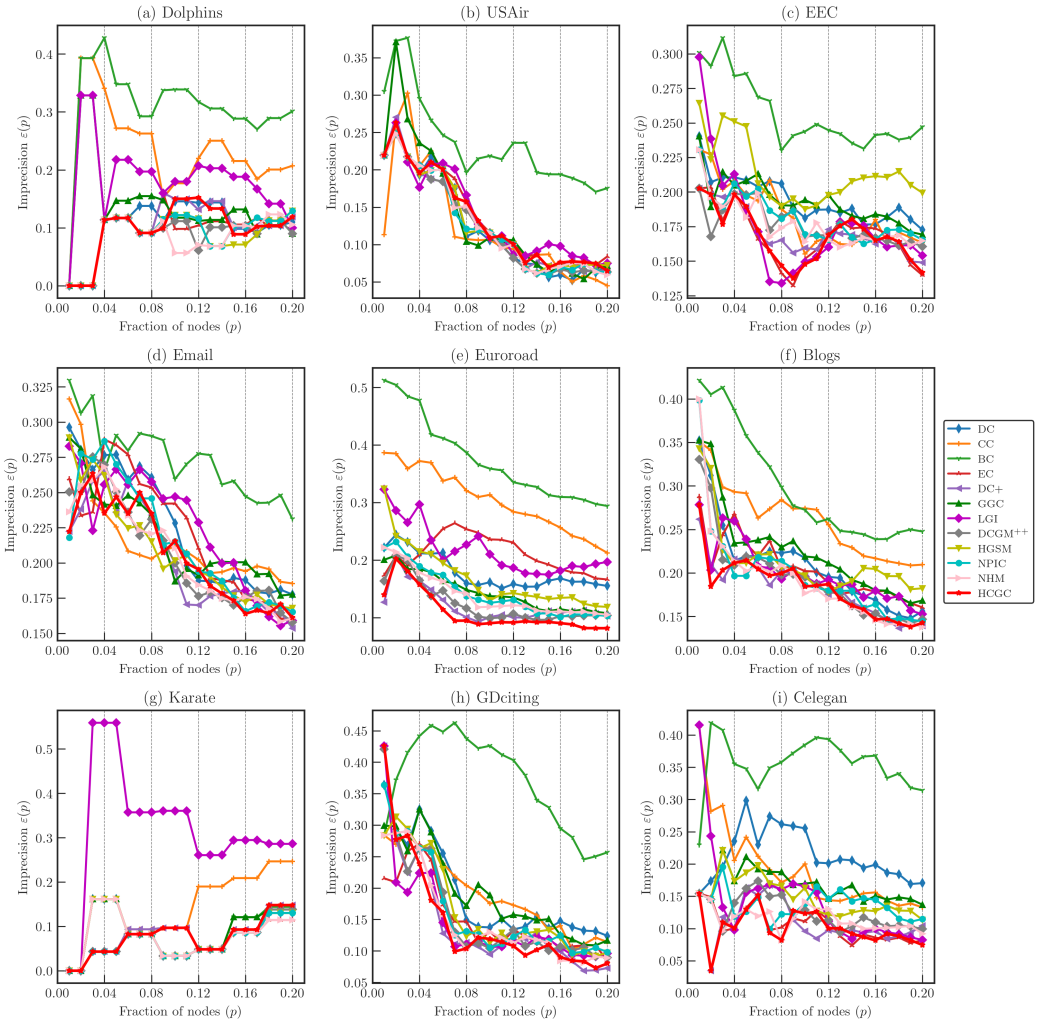


Figure 5. Comparison of imprecision function of different measures on public real networks dataset with $\beta = \beta_c$ and $R = 3$.

5.3.1. Centrality Measurement in Province Network

For the province network, measures across 77 nodes are applied to identify potential hubs. Each node represents a Thai province across six regions, including northern, northeastern (Isan), western, central, eastern, and southern. Our analysis specifically focuses on degree centrality (DC), closeness centrality (CC), betweenness centrality (BC), and eigenvector centrality (EC) as baselines for comparison with HCGC with a truncation radius of $R = 3$. Each method offers valuable insights into the relative importance of provinces within the network.

Table 15 demonstrates the top 10 nodes with their scores for each measure. The results from HCGC show that Lopburi province has the highest score, followed by Khon Kaen, Nakhon Ratchasima, Nakhon Sawan, and Phetchabun. These results include nodes from the top-ranked nodes of baseline measures, demonstrating that the influential nodes identified by our approach consider not only local information (like DC and EC) but also global information (like CC and BC). Moreover, our proposed method produces scores with high variance, which helps identify vital hubs more easily compared to traditional DC or CC, which often produce similar scores that are difficult to distinguish and rank. Additionally, the top 10 provinces come from diverse regions of Thailand, including central, northeastern, and northern.

Table 15. Top 10 provinces in the province network ranked by different measures.

Rank	DC		CC		BC		EC		HCGC			
	Node	Score	Node	Score	Node	Score	Node	Score	Node	Province	Region	Score
1	14	0.1184	6	0.2533	61	0.3211	6	0.2734	6	Lopburi	Central	217,858.6
2	28	0.1184	58	0.2476	55	0.3062	18	0.2454	28	Khon Kaen	North eastern	216,276.7
3	50	0.1184	47	0.2460	62	0.3046	47	0.2357	18	Nakhon Ratchasima	North eastern	200,406.2
4	18	0.1053	4	0.2452	69	0.2874	28	0.2286	47	Nakhon Sawan	Central	200,112.5
5	6	0.1053	50	0.2452	67	0.2293	54	0.2258	54	Phetchabun	Northern	191,681.5
6	47	0.1053	56	0.2428	50	0.2164	4	0.2182	50	Tak	Northern	172,389.4
7	40	0.0921	9	0.2397	28	0.1930	9	0.1897	4	Phra Nakhon Si Ayutthaya	Central	152,853.1
8	57	0.0921	48	0.2397	56	0.1929	57	0.1883	16	Nakhon Nayok	Central	133,072.0
9	54	0.0921	18	0.2397	58	0.1702	16	0.1863	58	Nakhon Pathom	Central	130,613.8
10	4	0.0921	54	0.2390	18	0.1637	3	0.1822	9	Saraburi	Central	127,022.9

Figure 6 presents nodes in the province network with their province labels and communities categorized by the Louvain algorithm. Larger node sizes represent higher HCGC scores. There are six communities in the network, which corresponds to the number of regions in Thailand.

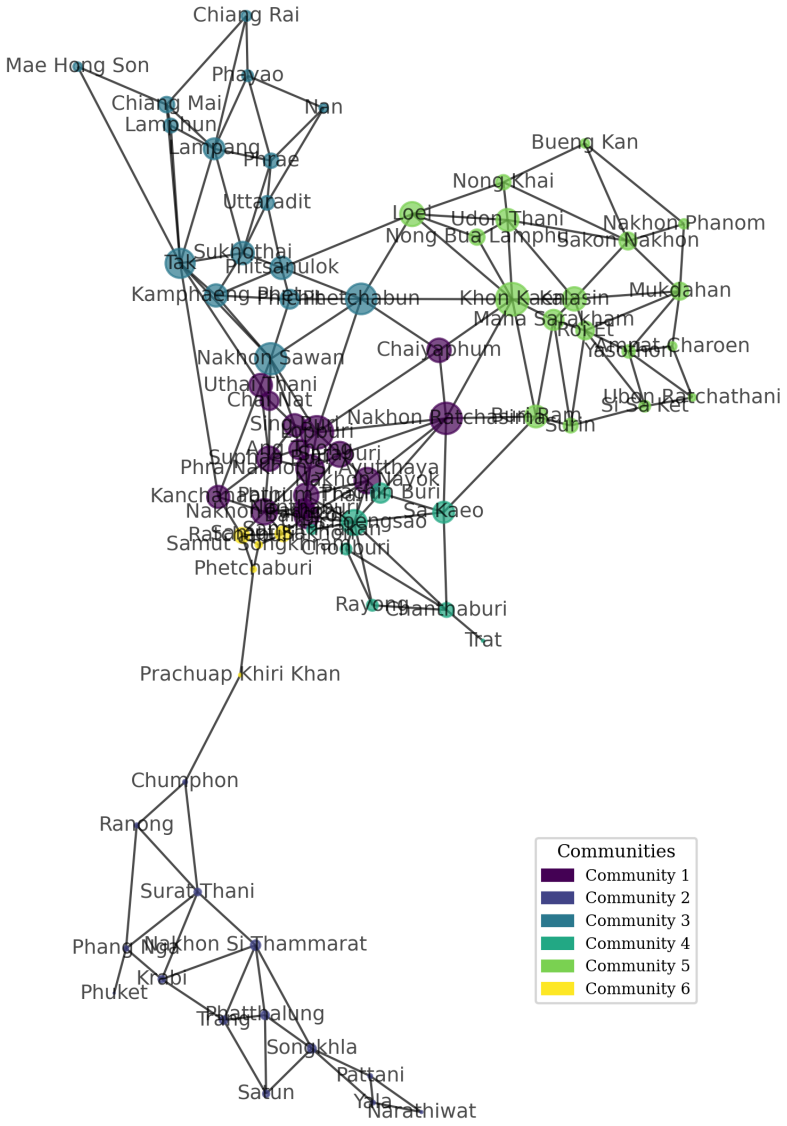


Figure 6. The nodes in province network with their province labels and communities.

Figure 7 illustrates the top 20% of nodes representing the vital hubs in the province network with their province labels and communities. The vital hubs exist in Communities 1, 3, 4, and 5. The results show that Lopburi province, which achieves the highest score using the HCGC method by connecting nodes from other communities, can be considered the most vital hub in this network.

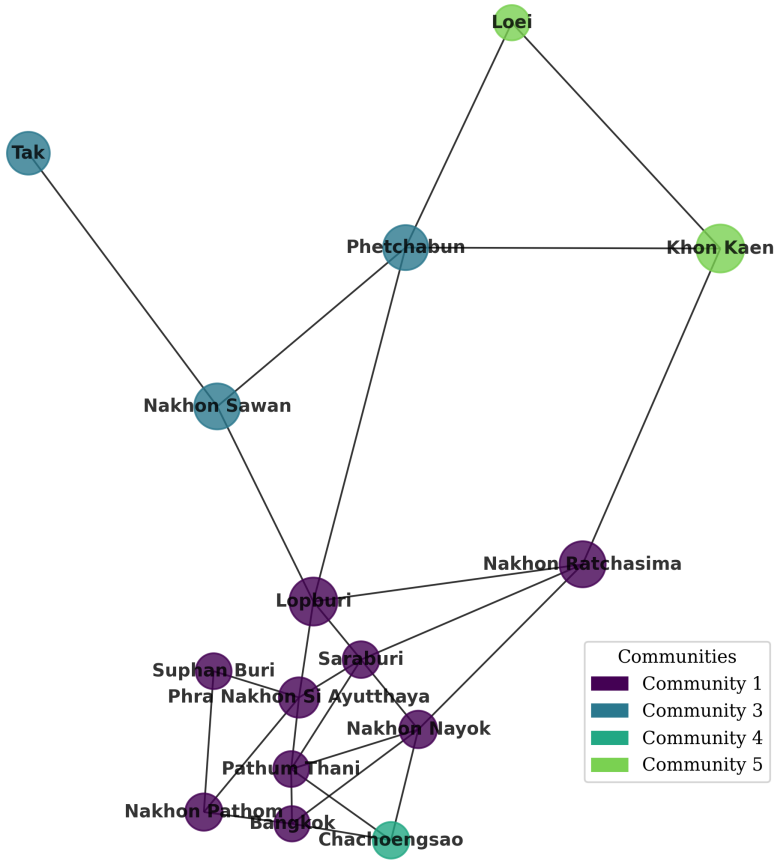


Figure 7. The nodes of top 20% in province network with their province labels and communities.

5.3.2. Centrality Measurement in District Network

For the district network, we analyze centrality measures across 928 Thai districts in the ThaiNet dataset. Similar to the centrality measurement in the province network, each node represents a district. DC, CC, BC, EC, and HCGC are used to calculate the scores for each node.

Table 16 describes the top 10 districts with scores of different measures. The results from HCGC show that Pathum Wan district in Bangkok province achieves the highest performance, followed by Thon Buri, Dusit, Sa Thon, and Bang Rak districts. All of these districts are located in Bangkok province. HCGC presents scores with high variance, which makes it easier to differentiate influential nodes from each other compared to DC.

Figure 8 presents nodes in the district network with their district labels and communities categorized by the Louvain algorithm. There are 32 communities in the network. The nodes have different sizes according to their HCGC scores. Higher scores correspond to larger node sizes.

Figure 9 illustrates the nodes in the first community. Due to the large number of communities, we select Community 1 as a representative example. The color bar on the right side shows the HCGC values. High-scoring nodes such as node 6 (Pathum Wan), node 14 (Thon Buri), and node 1 (Dusit) are labeled in yellow.

Table 16. Top 10 districts in the district network ranked by different measures.

Rank	DC		CC		BC		EC		HCGC			
	Node	Score	Node	Score	Node	Score	Node	Score	Node	District	Province	Score
1	14	0.0216	127	0.0861	776	0.2728	6	0.2537	6	Pathum Wan	Bangkok	1.110×10^8
2	1	0.0205	340	0.0854	777	0.2695	14	0.2503	14	Thon Buri	Bangkok	1.093×10^8
3	6	0.0194	101	0.0852	771	0.2681	1	0.2401	1	Dusit	Bangkok	1.076×10^8
4	0	0.0183	183	0.0851	770	0.2667	0	0.2337	27	Sa Thon	Bangkok	9.467×10^7
5	27	0.0183	705	0.0849	772	0.2653	12	0.2328	3	Bang Rak	Bangkok	9.421×10^7
6	32	0.0173	118	0.0847	773	0.2638	27	0.2317	0	Phra Nakhon	Bangkok	9.341×10^7
7	53	0.0173	203	0.0847	774	0.2624	3	0.2299	12	Samphanthawong	Bangkok	8.930×10^7
8	12	0.0173	709	0.0844	844	0.2567	17	0.2276	17	Khlong San	Bangkok	8.745×10^7
9	24	0.0173	215	0.0841	841	0.2531	7	0.2115	24	Bang Phlat	Bangkok	7.736×10^7
10	38	0.0173	216	0.0840	769	0.2479	30	0.1989	30	Bang Kho Laem	Bangkok	7.590×10^7

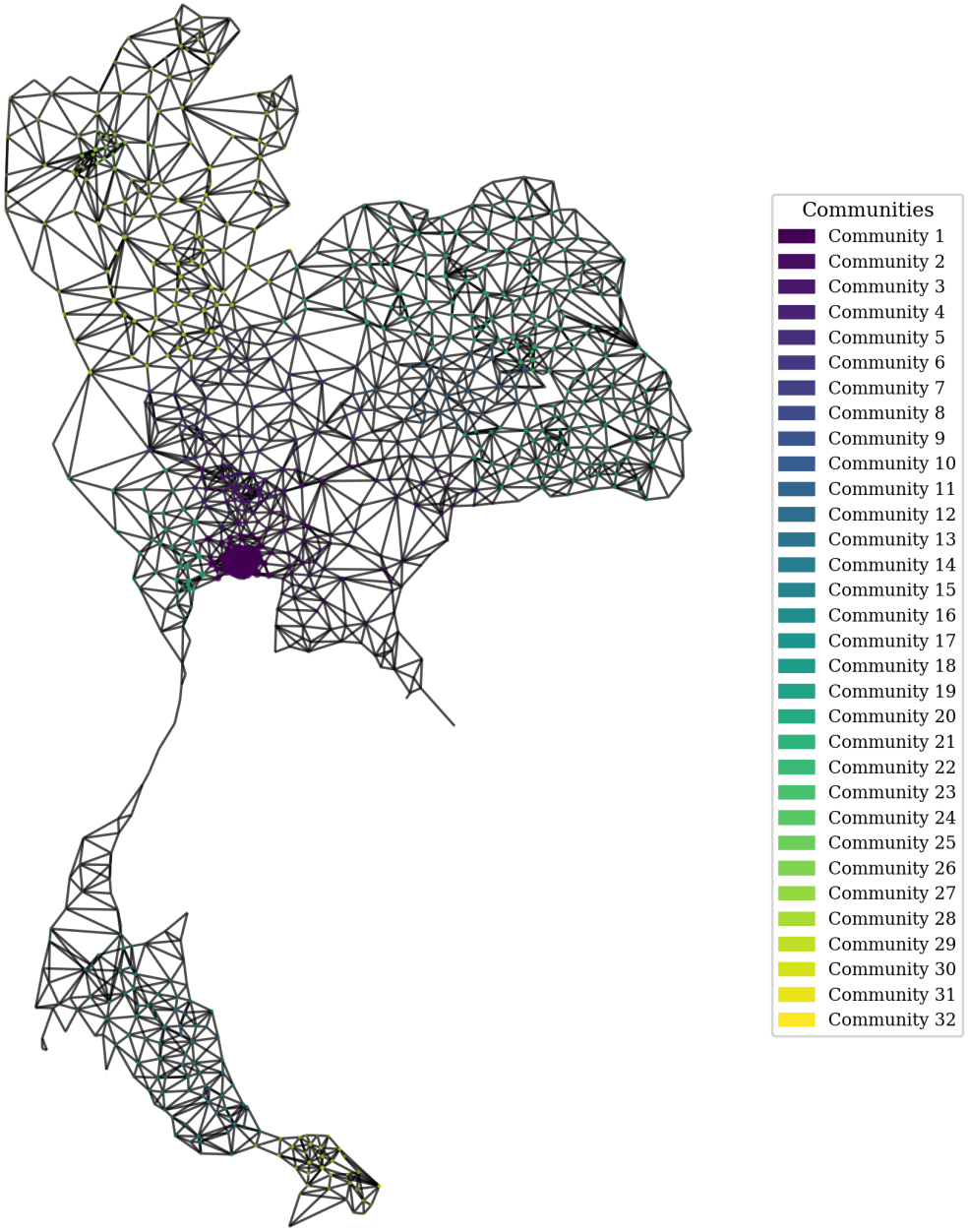


Figure 8. The nodes in the district network with their district labels and communities.

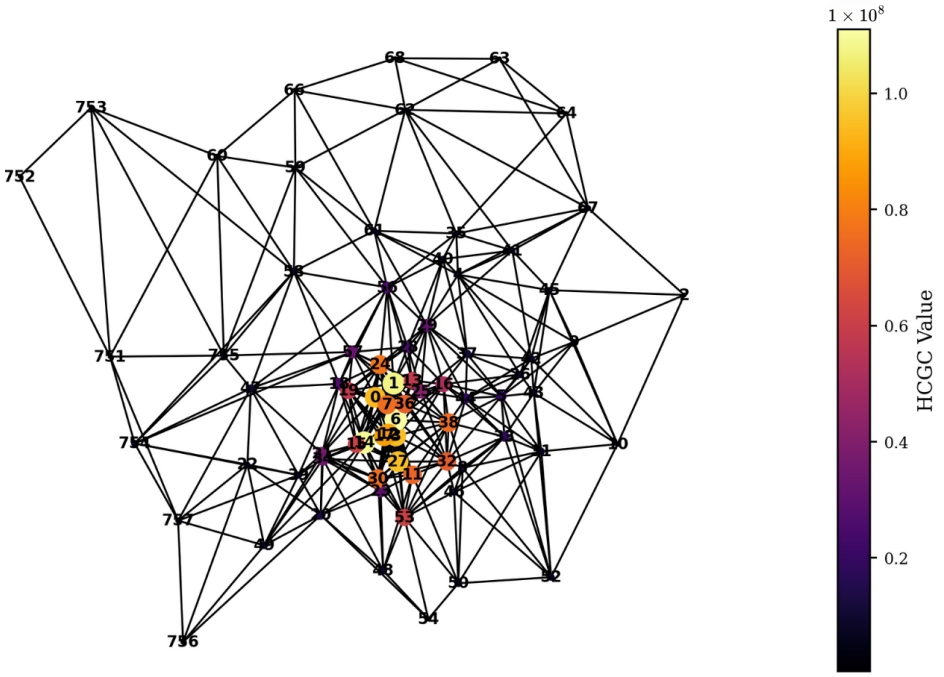


Figure 9. The nodes in the first community which contains the vital hubs in the network.

6. Conclusions

This paper proposes a novel approach called Hybrid Community-based Gravity Centrality (HCGC) for identifying hubs in transportation networks by employing local modularity, community detection, and gravity-based modeling. According to evaluation results on nine public real-world networks, HCGC outperforms existing approaches for identifying influential nodes in complex networks across different aspects. HCGC with a truncation radius $R = 2$ shows the high values of Kendall's correlation coefficient with SIR models on six out of nine networks. However, HCGC with a truncation radius $R = 3$ presents high values of the coefficient with SIR models on five out of nine networks compared to the traditional approaches. Additionally, HCGC with both truncation radii also achieves the best average imprecision function scores across all nine public real-world networks. This can be explained by the fact that our approach is outstanding in incorporating spreading influence that captures diffusion patterns in transportation networks.

In addition to the public networks dataset, we present the ThaiNet dataset consisting of two comprehensive regional transportation networks in Thailand including the province network and district network constructed by using an adjacency matrix of the nation's provinces and districts, respectively. It represents Thailand's transportation network across 77 provinces and 928 districts in Thailand. Once applying HCGC to this ThaiNet dataset, HCGC shows high variance scores compared to existing network centrality measures (such as DC or CC) that possibly produces the same scores for multiple nodes. This allows the nodes in the network to be more easily ranked and identified as hubs with accurate scores.

Based on the application of HCGC to the province network, provinces across six diverse regions can be considered hubs in the network. Lopburi is considered a hub in the central region with the top scores among different provinces in Thailand followed by Khon Kaen and Nakhon Ratchasima in the northeastern region. Phetchabun and Tak are two provinces in the northern region that achieve the highest scores in the region and are listed in the top ten provinces with the highest HCGC score. In addition, HCGC was applied to the district network. The districts that are listed in the top ten districts of high HCGC score represent the highly potential hubs in Thailand. All these districts are located in Bangkok province which is the capital of Thailand. The first rank district ordered

by HCGC is Pathum Wan followed by Thon Buri and Dusit, respectively. Furthermore, the communities for both networks are determined by the Louvain algorithm. There are six communities in the province network and thirty-two communities in the district network that can be categorized. These communities can be used to indicate vital nodes that have high potential to be hubs of Thailand and can visualize the size and location of each node in the map.

Our proposed method and application demonstration provide not only in-depth geographical and economic analysis but also a strong foundation for policy-making aimed at bolstering national connectivity and development. Moreover, the adaptable methodology proposed here can be applied to other geographical areas, enabling systematic data collection and analysis worldwide. As such, the datasets and methods developed through this study are invaluable resources that can drive informed decision-making and stimulate socio-economic progress across various regions.

Even though this study shows significant improvements over existing approaches with strong methodological and empirical contributions, there are limitations and opportunities for improvement. First, the transportation networks in the ThaiNet dataset were modeled as unweighted binary networks for simplicity, which may not reflect real-world systems that incorporate weighted factors such as distance, traffic volume, or speed. Second, processing large network datasets can be time-consuming due to the computational complexity of the HCGC measure. To address these limitations, future work will involve constructing weighted networks using actual road distances or travel times between districts, requiring substantial data collection and careful review of licensing requirements. This will enable a more realistic analysis of spatial interactions and proximity effects. Additionally, a more efficient approach for large-scale networks will be explored.

Author Contributions: Conceptualization, A.P., N.H., S.S. and W.T.; data curation, A.P. and W.T.; methodology, A.P. and W.T.; software, A.P., S.S. and W.T.; validation, W.T.; formal analysis, S.S. and W.T.; visualization, A.P., S.S. and W.T.; writing—original draft preparation, A.P., N.H., S.S. and W.T.; writing—review and editing, A.P., N.H., S.S. and W.T.; supervision, N.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Chiang Mai University Junior Research Fellowship Program.

Data Availability Statement: The datasets collected, compiled, and analyzed in this study, along with the code used to generate them, are publicly available and can be accessed through our GitHub repository: https://github.com/wtepsan/Adjacency-and-Distance-Matrix-of-Thailand, accessed on 1 March 2024. Additionally, the foundational public datasets used as the basis for generating the core dataset for this study are listed below:

- **Thailand District Boundaries:** Provided by the Geo-Informatics and Space Technology Development Agency, Thailand, and available at: https://opendata.onde.go.th/dataset/8-administrative-boundaries or https://github.com/wtepsan/Adjacency-and-Distance-Matrix-of-Thailand/tree/main/data_foundation_public_source (accessed on 1 March 2024).
- **Thailand District Lists:** Available at either https://data.go.th/dataset/view_district or https://data.go.th/th/dataset/item_f9a9a9dd-d23d-4b86-89ae-e34820d4f3dc, accessed on 1 March 2024.
- **Network Datasets for Model Evaluation:** Selected from public open datasets, including NETWORKREPOSITORY (https://networkrepository.com/networks.php) (accessed on 14 February 2025) and KONECT (http://konect.cc/networks/) (accessed on 14 February 2025).

Acknowledgments: In addition to the support from the Chiang Mai University Junior Research Fellowship Program, we would like to express our gratitude to the Erawan HPC Project, Information

Technology Service Center (ITSC), Chiang Mai University, Chiang Mai, Thailand, for providing the computational resources essential to this research.

Conflicts of Interest: The author declare no conflicts of interest.

References

1. GIS and Network Analysis. In *Spatial Analysis and GeoComputation: Selected Essays*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 43–60. [CrossRef]
2. Huang, B.; Cova, T.; Tsou, M.H.; Bareth, G.; Song, C.; Song, Y.; Cao, K.; Silva, E. *Comprehensive Geographic Information Systems*; Elsevier: Amsterdam, The Netherlands, 2018.
3. Zinoviev, D. *Complex Network Analysis in Python: Recognize–Construct–Visualize–Analyze–Interpret*; The Pragmatic Bookshelf: Raleigh, NC, USA, 2018.
4. Scully, P. Thailand Province Border-Adjacency Mappings. GitHub Repository. 2021. Available online: https://github.com/pmdscully/thailand_province_border_adjacency (accessed on 1 March 2024).
5. Scully, P. Thailand Province Border Adjacency Dataset/Code. 2021. Available online: https://petescully.co.uk/2021/02/08/thailand-province-border-adjacency-dataset/ (accessed on 1 March 2024).
6. Ford, D.; Kaufman, J.; Eiron, I. An Extensible Spatial and Temporal Epidemiological Modelling System. *Int. J. Health Geogr.* **2006**, *5*, 4. [CrossRef] [PubMed]
7. Chinpong, K.; Thavornwattana, K.; Armatrmontree, P.; Chienwichai, P.; Lawpoolsri, S.; Silachamroon, U.; Maude, R.J.; Rotejanaprasert, C. Spatiotemporal Epidemiology of Tuberculosis in Thailand from 2011 to 2020. *Biology* **2022**, *11*, 755. [CrossRef] [PubMed]
8. Martínez Márquez, R.A.; Patanè, G. Graph Node Scoring for the Analysis and Visualisation of Mobility Networks and Data. *Urban Sci.* **2024**, *8*, 155. [CrossRef]
9. Mansourihanis, O.; Maghsoodi Tilaki, M.J.; Yousefian, S.; Zaroujtaghi, A. A Computational Geospatial Approach to Assessing Land-Use Compatibility in Urban Planning. *Land* **2023**, *12*, 2083. [CrossRef]
10. Lü, L.; Chen, D.; Ren, X.L.; Zhang, Q.M.; Zhang, Y.C.; Zhou, T. Vital nodes identification in complex networks. *Phys. Rep.* **2016**, *650*, 1–63. [CrossRef]
11. Chakravarthy, T.S.; Lokesh, S. *A Comprehensive Review of Influence Node Identification in Complex Networks*; ACM: New York, NY, USA, 2022; pp. 11–26.
12. Wang, P.; Zeng, C.; Song, Y.; Guo, L.; Liu, W.; Zhang, W. The Spatial Effect of Administrative Division on Land-Use Intensity. *Land* **2021**, *10*, 543. [CrossRef]
13. Xu, R.; Chen, Z.; Li, F.; Zhou, C. Identification of Urban Functional Zones Based on POI Density and Marginalized Graph Autoencoder. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 343. [CrossRef]
14. Kang, C.; Jiang, Z.; Liu, Y. Measuring hub locations in time-evolving spatial interaction networks based on explicit spatiotemporal coupling and group centrality. *Int. J. Geogr. Inf. Sci.* **2022**, *36*, 360–381. [CrossRef]
15. Bellingeri, M.; Bevacqua, D.; Scotognella, F.; Zhe-Ming, L.; Cassi, D. Efficacy of local attack strategies on the Beijing road complex weighted network. *Phys. A Stat. Mech. Its Appl.* **2018**, *510*, 316–328. [CrossRef]
16. Tian, T.; Cheng, Y.; Liang, Y.; Ma, C.; Chen, K.; Hu, X. Hub Node Identification in Urban Rail Transit Network Evolution Using a Ridership-Weighted Network. *Transp. Res. Rec.* **2024**, *2678*, 549–569. [CrossRef]
17. Yuan, G.; Sun, L.; Kong, D.; Bai, Z.; Shao, J. Supernetwork Perspective on Studying the Method of Identifying the Important Urban Transport Hub. *J. Highw. Transp. Res. Dev. (Engl. Ed.)* **2023**, *17*, 82–91. [CrossRef]
18. Eydi, A.; Shirinbayan, P. Multi-modal and multi-product hierarchical hub location problem with fuzzy demands. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106282. [CrossRef]
19. Yıldız, B.; Yaman, H.; Karaşan, O.E. Hub Location, Routing, and Route Dimensioning: Strategic and Tactical Intermodal Transportation Hub Network Design. *Transp. Sci.* **2021**, *55*, 1351–1369. [CrossRef]
20. Landherr, A.; Friedl, B.; Heidemann, J. A critical review of centrality measures in social networks. *Bus. Inf. Syst. Eng.* **2010**, *2*, 371–385. [CrossRef]
21. Rodrigues, F.A. Network centrality: An introduction. In *A Mathematical Modeling Approach from Nonlinear Dynamics to Complex Systems*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 177–196.
22. Malang, K.C.; Wang, S.; Phaphuangwittayakul, A.; Lv, Y.; Yuan, H.; Zhang, X. Identifying influential nodes of global terrorism network: A comparison for skeleton network extraction. *Phys. A Stat. Mech. Its Appl.* **2020**, *545*, 123769. [CrossRef]
23. Ait Rai, K.; Machkour, M.; Antari, J. Influential nodes identification in complex networks: A comprehensive literature review. *Beni-Suef Univ. J. Basic Appl. Sci.* **2023**, *12*, 18. [CrossRef]
24. Lai, Q.; Zhang, H.H. Analysis of identification methods of key nodes in transportation network. *Chin. Phys. B* **2022**, *31*, 068905. [CrossRef]

25. Rhoads, D.; Rames, C.; Solé-Ribalta, A.; González, M.C.; Szell, M.; Borge-Holthoefer, J. Sidewalk networks: Review and outlook. *Comput. Environ. Urban Syst.* **2023**, *106*, 102031. [CrossRef]
26. Li, H.; Shang, Q.; Deng, Y. A generalized gravity model for influential spreaders identification in complex networks. *Chaos Solitons Fractals* **2021**, *143*, 110456. [CrossRef]
27. Li, Z.; Ren, T.; Ma, X.; Liu, S.; Zhang, Y.; Zhou, T. Identifying influential spreaders by gravity model. *Sci. Rep.* **2019**, *9*, 8387. [CrossRef]
28. Mukhtar, M.F.; Abal Abas, Z.; Baharuddin, A.S.; Norizan, M.N.; Fakhruddin, W.F.W.W.; Minato, W.; Rasib, A.H.A.; Abidin, Z.Z.; Rahman, A.F.N.A.; Anuar, S.H.H. Integrating local and global information to identify influential nodes in complex networks. *Sci. Rep.* **2023**, *13*, 11411. [CrossRef] [PubMed]
29. Chen, D.; Su, H. Identification of influential nodes in complex networks with degree and average neighbor degree. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2023**, *13*, 734–742. [CrossRef]
30. Souravlas, S.; Sifaleras, A.; Tsintogianni, M.; Katsavounis, S. A classification of community detection methods in social networks: A survey. *Int. J. Gen. Syst.* **2021**, *50*, 63–91. [CrossRef]
31. Miller, J.; Ting, T. EoN (Epidemics on Networks): A fast, flexible Python package for simulation, analytic approximation, and analysis of epidemics on networks. *J. Open Source Softw.* **2019**, *4*, 1731. [CrossRef]
32. Kendall, M.G. A new measure of rank correlation. *Biometrika* **1938**, *30*, 81–93. [CrossRef]
33. Esfandiari, S.; Fakhrahmad, S.M. Identifying influential nodes in complex networks by adjusted feature contributions and neighborhood impact. *J. Supercomput.* **2025**, *81*, 1–39. [CrossRef]
34. Qiu, L.; Zhang, J.; Tian, X. Ranking influential nodes in complex networks based on local and global structures. *Appl. Intell.* **2021**, *51*, 4394–4407. [CrossRef]
35. Ullah, A.; Sheng, J.; Wang, B.; Din, S.U.; Khan, N. Leveraging neighborhood and path information for influential spreaders recognition in complex networks. *J. Intell. Inf. Syst.* **2024**, *62*, 377–401. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.