

Received July 1, 2021, accepted July 21, 2021, date of publication July 28, 2021, date of current version August 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3101176

On the Utilization of Shortest Paths in Complex Networks

HEND ALRASHEED 

Department of Information Technology, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia

e-mail: halrasheed@ksu.edu.sa

This work was supported by the Research and Development Program (Research Pooling Initiative), Ministry of Education, Riyadh, Saudi Arabia.

ABSTRACT Considerable effort has been devoted to the study of network structures and connectivity patterns and their influence on network dynamics. A widely used assumption in network analysis models is that traffic follows the shortest paths connecting pairs of nonneighboring vertices. For example, graph centrality measures, community extraction algorithms, and core-periphery detection algorithms use this assumption. However, this is a very restricted perspective and can be misleading as a consequence of its focus on shortest path communications. In this work, we study the utilization of shortest paths in complex networks in different data dissemination scenarios. We also explore whether there are general properties that can make networks utilize shortest paths more effectively. By conducting simulations on a set of real-world and artificial networks, we show that the utilization of shortest paths in complex networks may not be as common as assumed. This implies that longer paths can be as important (in some cases) as the shortest paths. Our results show that at least two factors clearly influence shortest path utilization in a network: the structure of the network and the data dissemination algorithm. We also find that the type of a network is not a good indicator of its shortest path utilization.

INDEX TERMS Influence maximization (IM) model, network analysis, network distance properties, network structure, shortest paths, small-world phenomenon, susceptible–infectious–recovered (SIR) model, vertex centrality.


I. INTRODUCTION

Networks describe the bidirectional relationships among entities in systems, for instance, social systems, biological systems, and the World Wide Web. In its simplest form, a network is a set of entities and their connections. Combining network science with the mathematical perspective of graph theory provides a rich toolbox for analyzing and understanding various complex phenomena of networks.

Mathematically, graphs are used to represent real-world networks. A graph is represented by a set of vertices (entities or agents) and a set of edges (relationships or connections between vertices). Depending on the system, a graph can be weighted or unweighted. The weight of an edge between two vertices captures the notion of the cost required to travel through that edge. If no edge exists between a pair of vertices, then the two vertices may be connected by a path that consists of a sequence of intermediate vertices. The distance of a path

is the sum of the weights of all the edges on that path, i.e., the total cost. A path is considered a shortest path (also called a geodesic) if it has minimum cost.

Significant effort has been devoted to the study of network structures and connectivity patterns and their influence on a network's dynamics [1]–[11]. A widely used assumption in network analysis models is that traffic follows the shortest paths connecting pairs of nonneighboring vertices [11], [12]. Examples using this assumption include graph centrality measures, which rank vertices according to their importance based on some criteria [6], [13], [14]; community extraction algorithms, which identify several groups of interconnected sets [15]–[17]; and core-periphery detection algorithms, which partition vertices into a densely connected core and a sparsely connected periphery [18]–[20]. These measures and models are based on the assumption that information is passed from one vertex to another only along the shortest paths linking them. However, this perspective is very restricted and can be misleading as a consequence of its focus on shortest path communications [21], [22].

The associate editor coordinating the review of this manuscript and approving it for publication was Weiguo Xia .

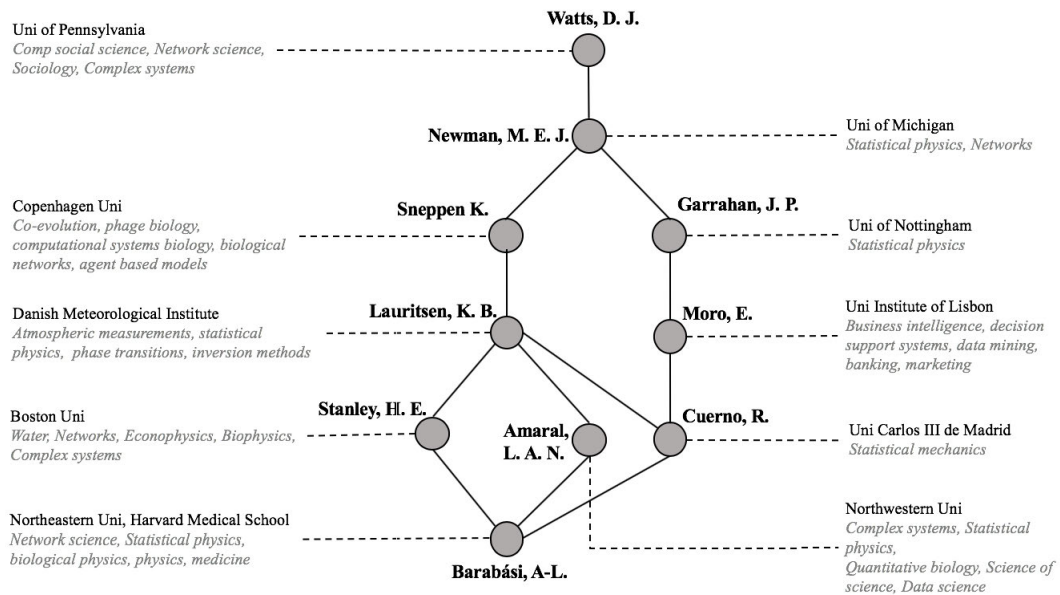


FIGURE 1. Collaboration network of physics papers highlighting the shortest paths from Newman, M. E. J. to Barabási, A.-L. The figure was adapted from [27].

Real-life examples that contradict the assumption of shortest path navigation can be observed in several network applications. For example, the small-world phenomenon,¹ which has been observed in a wide range of network types and forms the basis of many network analysis models, was inspired by an experiment that received several methodological criticisms [23]–[25]. In 1967, Milgram [26] performed an experiment to examine the average path lengths of social networks in the United States. He asked a set of participants to deliver a letter to a target address by advancing it to an acquaintance, and found that the average path length was six. The experiment had a high noncompletion rate [23] and the relevance of indirect contact chains of different degrees of separation is questionable [25].

The second example is related to relationship formation through shortest paths. Newman analyzed scientists' collaboration networks using bibliometrics information. He found that scientists were more likely to collaborate if they had a common collaborator [27]. However, the role that intermediate scientists played in connecting other noncollaborating scientists was not clear, i.e., Did the existence of shortest paths between nonneighboring author pairs suggest future collaborations? In Fig. 1, which appeared in [27], although the distance between Newman, M. E. J. and Moro, E. was two, these scientists have never collaborated. Similarly, with respect to publishing together, Watts, D. J. and Sneppen K., Garrahan J. P. and Lauritsen, K. B., and Stanley H. E. and

¹The small-world property refers to the fact that in most real-world networks, the typical geodesic distance is short, in particular, when compared with the network size [6]. For example, people are linked through short chains of acquaintances in a social network.

Garrahan J. P., with distances of 2, 3, and 4, respectively, have never published together.²

Here, in an attempt to investigate the shortest path communication assumption, we study the utilization of shortest paths in networks in different data dissemination scenarios. Moreover, we explore whether there are general properties that can make networks utilize shortest paths more effectively. By conducting a number of simulations on a set of real-world and artificial networks, we show that the utilization of shortest paths in complex networks may not be as common as assumed. This implies that longer paths can be as important (in some cases) as the shortest paths. Our results show that at least two factors clearly influence shortest path utilization in a network: the network structure and the data dissemination algorithm. We also find that the network type is not a good indicator of its shortest path utilization.

Research has increasingly focused on how the structure of networks affects their dynamics and evolution. For instance, network structure was shown to have a significant impact on spreading processes; however, a specific characterization of the interplay involved is yet to be presented. Our results are expected to have profound implications for how we understand diffusion dynamics and relationship formation in complex networks. Such an understanding could, for instance, increase preparedness against emerging diseases for which limited epidemiological data are available. It could also guide the creation of network evolution models that provide a more realistic characterization of human communication.

²The profiles of all listed authors were examined up to the publication date of this paper.

TABLE 1. This table summarizes notation used frequently throughout this work.

Parameter	Definition
$G = (V, E)$	Graph G with vertex set V and edge set E .
$size(G)$	Graph size; $size(G) = V + E $.
$\rho(u, v)$	A shortest path connecting two vertices u and v .
$d(u, v)$	Distance between two vertices u and v , defined as the length of the shortest path that connects u and v .
$l(G)$	Graph average shortest path length; $l(G) = \frac{1}{n(n-1)} \sum_{u \in V} \sum_{v \in V} d(u, v)$.
$ecc(u)$	Eccentricity of vertex u ; $ecc(u) = \max_{v \in V} \{d(u, v)\}$.
$diam(G)$	Graph diameter; $diam(G) = \max_{u, v \in V} \{d(u, v)\}$.
$rad(G)$	Graph radius; $rad(G) = \min_{u \in V} \{ecc(u)\}$.
$C(G)$	Graph center; $C(G) = \{u \in V : ecc(u) = rad(G)\}$.
$N(u)$	Neighborhood of vertex u ; $N(u) = \{v \in V : uv \in E\}$.
$degree(u)$	Number of vertices in vertex u 's neighborhood (undirected graph); $degree(u) = N(u) $.

The rest of this paper is organized as follows. In Section II, we briefly introduce preliminary concepts that will be used throughout this work. Section III presents relevant related work. In Section IV, we describe the network datasets used in this work. Sections V, VI, and VII present the method, results, and a discussion, respectively. Finally, Section VIII concludes this work.

II. PRELIMINARIES

In this section, we review a number of relevant graph theory concepts and definitions. A summary of all notation is provided in Table 1.

A graph $G = (V, E)$ is an ordered pair (V, E) , where V is the set of vertices and $E \subseteq V^2$ is the set of edges. We use $|V| = n$ to denote the number of vertices and $|E| = m$ to denote the number of edges in G . We define the size of the graph, denoted by $size(G)$, as the sum of the number of vertices and the number of edges, i.e., $size(G) = |V| + |E|$.

Depending on the system being modeled, a graph can be directed or undirected. Similarly, a graph can be weighted or unweighted. When G is weighted, each edge between two vertices u and v is assigned a weight value e_{uv} . When the graph is unweighted, the value of the weight will be binary (1 and 0 are often used to indicate the existence or the lack of an edge, respectively). When an edge exists between two vertices $u, v \in V$, then the two vertices are called *neighbors*. A pair of nonneighboring vertices u and v may be connected by a path that includes the sequence of all intermediate vertices between u and v . All graphs in this work are undirected and unweighted.

A path P of length k between two vertices u and v is a sequence of adjacent vertices u_0, u_1, \dots, u_k , where $u_0 = u$ and $u_k = v$, and a shortest path (geodesic) from u to v , denoted by $\rho(u, v)$, is a path that minimizes the length of this path. Note that multiple shortest paths may exist between a pair of vertices. The distance of a path $\rho(u, v)$, denoted by $d(u, v)$, is the sum of the weights of all the edges on that path, i.e., the total cost. A path $\rho(u, v)$ is considered a shortest path if it is a path with minimum cost.

Three key distance properties are related to the structure of a graph: its average path length, diameter, and

small-worldness. The *average path length* $l(G)$ represents the average shortest path length averaged over all pairs of vertices. That is, $l(G) = \frac{1}{n(n-1)} \sum_{u \in V} \sum_{v \in V} d(u, v)$. The *diameter* of a given graph G is the length of the longest shortest path between any pair of vertices u and v in the graph. That is, $diam(G) = \max d(u, v)$.

The *small-world* property refers to the fact that in most real-world networks, the typical geodesic distance is short, in particular, when compared with the network size [6]. A graph is said to exhibit the small-world property if its diameter is bounded by the logarithm of its size ($diam(G) \leq \log_2(size(G))$) [28].

The *eccentricity* of a vertex $u \in V$ ($ecc(u)$) is the distance between u and a vertex farthest from u , i.e., $ecc(u) = \max_{v \in V} \{d(u, v)\}$. Vertex eccentricity is a measure of how close a vertex is to every other vertex in the network. The value of the minimum eccentricity is known as the *radius* $rad(G)$ of the graph: $rad(G) = \min_{u \in V} \{ecc(u)\}$. The *center* of a graph $C(G)$ constitutes all vertices with minimum eccentricity ($C(G) = \{u \in V : ecc(u) = rad(G)\}$).

III. RELATED WORK

Many studies have examined the relationship dynamics in online social networks as an important aspect of their evolution. Wilson *et al.* [29] explored the relationship between social links and real user interactions. Studying user interactions in the Facebook network, the authors observed lower levels of small-world properties. That is, user interactions were more frequent between users who were not directly connected. In [30], the authors explored the connection between a network's structural properties and its dynamics. Specifically, they investigated the relationship between the shortest paths and the lengths of spreading paths. Using stochastic simulations and vertex sampling, they concluded that the spreading paths and shortest paths in complex networks coincided to a great extent.

The importance of indirect relationships in an individual's social network has been investigated. New link formation and information diffusion based on indirect ties (friends of friends) in social networks were investigated in [31]. The authors concluded that indirect ties could be used to predict

TABLE 2. Network datasets.

Dataset	No. of vertices	No. of edges	Density	Avg. degree	Diameter	Avg. path length	Small-worldness
KarateClub	34	78	0.14	4.6	5	2.4	1.8
Email	1133	5451	0.01	9.6	8	3.6	4.7
Facebook	4039	88234	0.01	43.7	8	3.7	8.5
DutchElite	3621	4310	0.001	2.4	22	8.6	-9
Watts–Strogatz(0.3)	2500	5000	0.002	4.0	12	6.4	0.9
Watts–Strogatz(0.5)	2500	5000	0.002	4	11	6.1	1.9
Watts–Strogatz(0.7)	2500	7491	0.002	6	10	3.8	3.3
Barabasi–Albert(3)	2500	7491	0.002	6	6	3.8	7.3
Barabasi–Albert(5)	2500	12475	0.004	10	5	3.3	8.9
Barabasi–Albert(7)	2500	17451	0.006	14	5	3	9.3
Erdos–Renyi(1.6)	1582	1688	0.001	2.1	35	14.3	-23.3
Erdos–Renyi(2)	1993	2418	0.001	2.4	25	10.1	-12.9
Erdos–Renyi(8)	2498	10026	0.003	8	7	4	6.6
PowerLaw(1.8)	2236	4122	0.001	3.7	11	4.5	1.6
PowerLaw(2)	1199	1304	0.002	2.2	23	8.5	-11.7
PowerLaw(2.7)	1761	2042	0.001	2.3	21	7.4	-9.1

the formation of new edges and information diffusion paths. Indirect relationships have been observed to influence the spread of obesity [32] and happiness [33] among individuals in a social network.

The utilization of shortest paths among commuters in road networks has been the subject of much research. Despite the widely accepted assumption that people follow shortest paths, recent findings show that commuters do not follow those paths most of the time. For example, Zhu and Levinson [12] studied day-to-day route choice behavior and identified a gap between the shortest paths and observed route decisions. They concluded that about two-thirds of commuters did not use the path with the shortest travel time, and no commuters followed the path with the shortest travel distance unless it coincided with the one with the shortest travel time. Similar results were reported in [34]. Tang and Levinson [35] evaluated routes followed by residents within the Minneapolis–Saint Paul area and found that most commuters used paths longer than the shortest paths. Moreover, they found that longer-distance trips tended to deviate more from the shortest path compared with short-distance trips.

Using real-life examples, the authors of [36] showed that shortest paths may not be the best possible paths. In [37], the shortest paths and empirical paths in multiple real-world networks were compared. The authors defined empirical paths as those determined by measurements, for example, traceroutes over the Internet network and path estimations in the brain network. They observed that empirical paths were 10–30% longer on average than the shortest paths in all networks. They explained this observation by the existence of internal network logic (in the form of various hierarchies) that affects the structure of the paths.

In [38], the interplay between network structure and dynamics was investigated using multivariate networks. They focused on network transitivity because of its role in the redundancy in the computation of shortest paths.

IV. DATASETS

In our investigation, it was important to use network datasets that exhibited multiple different structural features.

Therefore, we based our analysis on a set of real-world and artificial network datasets with different identifiable structural properties. A summary of the basic structural properties of each network dataset is given in Table 2 (the last three columns show the three distance properties of each network: diameter, average path length, and small-worldness). The distance matrix of each network is shown in Fig. 2.

A. REAL-WORLD NETWORKS

In most real-world networks, the network diameter and its average path length are small compared with the network's size. This is known as the small-world property [5], [39]. Most real-world networks are also scale-free. A network is scale-free when its vertex degrees exhibit a power-law distribution (i.e., the majority of vertices have small degrees and only very few vertices have higher degree) [3], [40].

We examined four widely used and publicly available social networks. The KarateClub network [41] captures the social ties between the members of a university karate club. The Email network [42] represents the email interchanges between members of the University of Rovira i Virgili, Tarragona. The DutchElite network [43] is a network dataset of the administrative elite in The Netherlands, where vertices represent individuals and organizations that are most important to the Dutch government (a two-mode network), and an edge connects two vertices if the individual vertex belongs to the organization vertex. The Facebook network [44] represents ego networks (the network of friendship among a user's friends) of 10 individuals; two vertices (users) are connected if they are Facebook friends.

B. WATTS–STROGATZ NETWORKS

The Watts–Strogatz network generation model [2] creates small-world networks with small average path lengths and high clustering coefficients. To generate such a network, a ring over n vertices is first created. Second, each vertex is connected with its k nearest neighbors. Then, some edges are rewired to create shortcuts with probability p . We generated three Watts–Strogatz networks each with 2500 vertices and with rewiring probabilities 0.3, 0.5, and 0.7, respectively.

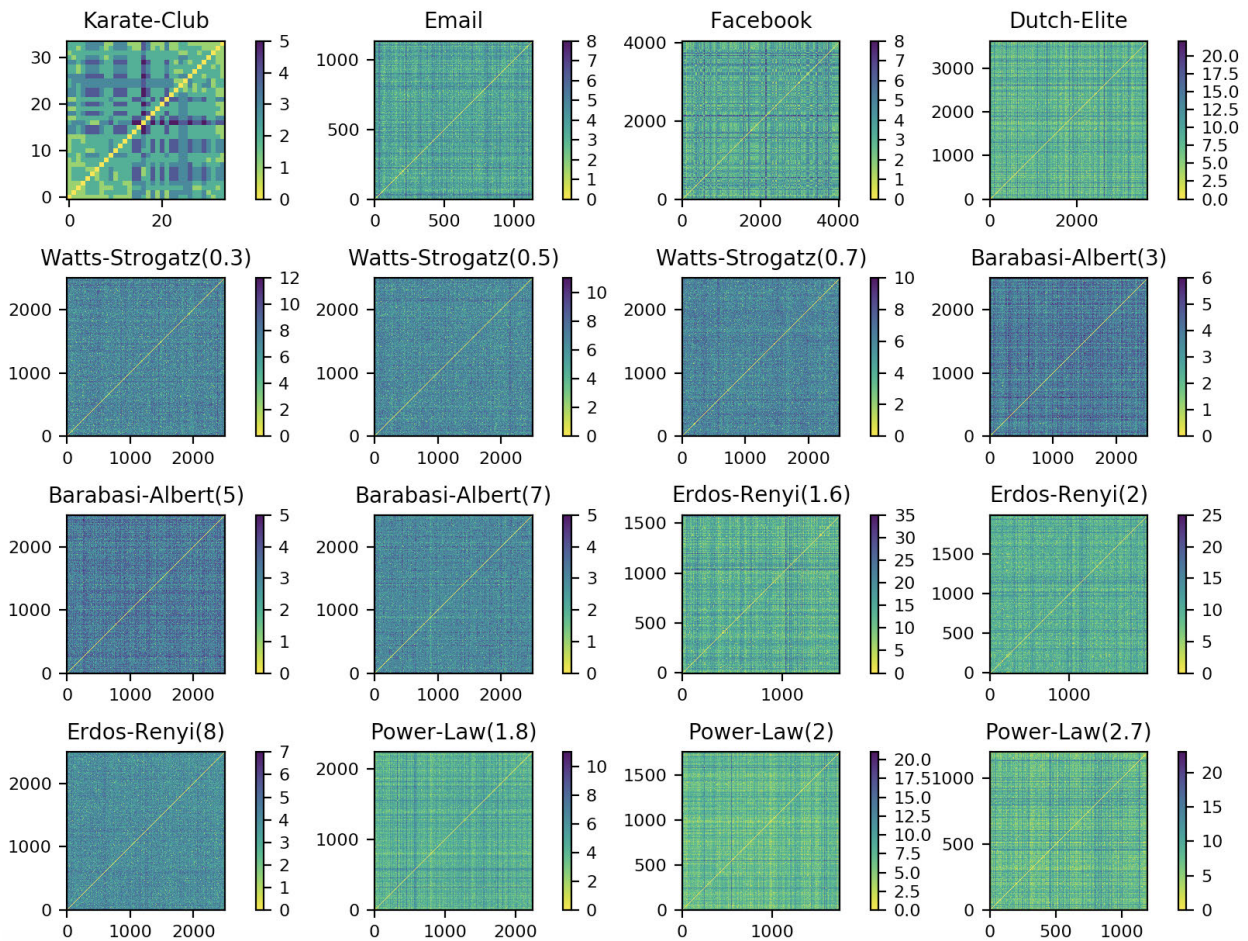


FIGURE 2. Network distance matrices.

C. BARABASI-ALBERT NETWORKS

The Barabasi-Albert model [45] generates scale-free networks with power-law degree distributions. A Barabasi-Albert network with n vertices is generated by adding new vertices one at a time and attaching each new incoming vertex to m existing vertices with high degrees. We generated three Barabasi-Albert networks, each with 2500 vertices and with m values of 3, 5, and 7, where m is the number of existing vertices to which a new vertex will attach.

D. ERDŐS-RÉNYI NETWORKS

In an Erdős-Rényi graph with n vertices, each pair of vertices is independently connected with probability p . Smaller values of p ($1/n < p < \log(n)/n$) result in very sparse graphs. By contrast, larger p values yield dense graphs with very small diameters. We generated three Erdős-Rényi graphs, each with 2500 vertices and with p values of $1.6/n$, $2/n$, and $8/n$.

E. POWER-LAW NETWORKS

In a power-law graph, the vertex degrees follow (or approximate) a power-law distribution. We generated power-law networks based on a variation of the Aiello-Chung-Lu

model [46], [47]. The degree sequence of each generated graph was determined by a power-law with exponent β , where β is the power parameter. Smaller β values ($\beta < 2$) generate power-law graphs with cores that are denser and have smaller diameters, compared to power-law graphs generated with higher β values [4]. Each power-law graph in the network dataset PowerLaw(β) had 2500 vertices and a value $\beta \in \{2.7, 2, 1.8\}$.

V. METHOD

This study investigated the characteristics of paths followed by various well-known data dissemination algorithms. The goal was to evaluate the utilization of shortest paths in complex networks and discover the topological properties that may influence this utilization. We investigated two algorithms: the SIR (Susceptible-Infectious-Recovered) infection-spreading model, and multiple implementations of the IM (Influence Maximization) model. All network datasets and code used in this section are available at https://github.com/halrashe/Shortest_Paths_Utilization.

A. SIR INFECTION-SPREADING MODEL

We first used the SIR infection-spreading model to examine the utilization of shortest paths in complex networks. In a

network system, the SIR infection-spreading model describes the discrete-time dynamics of an infection in a closed population. In such a system, vertices are partitioned into three compartments [48]: (1) susceptible, S , which represents the set of vertices susceptible to the disease; (2) infectious, I , which represents the set of vertices that have been infected and are able to spread the disease to susceptible vertices; and (3) recovered, R , which represents vertices that have recovered and cannot be infected again. At each time step, a susceptible vertex with an infected neighbor becomes infectious with probability p .

We ran the SIR infection-spreading model for each network in our network datasets as follows. Given a graph $G = (V, E)$ and a distinguished seed vertex s , we ran the SIR infection-spreading model for n time steps. We defined two attributes for each vertex $v \in V$ as follows.

(1) $seed(v)$ denotes the seed vertex that passed the infection to vertex v (directly or indirectly): direct infection occurs when a susceptible vertex v is a neighbor of seed vertex s , and indirect infection occurs when a susceptible vertex v is not a neighbor of seed vertex s .

(2) $\pi(v)$ denotes the number of infected vertices on the path from s to v (i.e., the number of vertices that are infected before passing the infection to vertex v); if v is susceptible at the end of the simulation, then both $seed(v)$ and $\pi(v)$ will be equal to ϕ .

At the end of each implementation, the total number of utilized shortest paths was computed as follows. Let $d(s, v)$ be the distance of a shortest path connecting vertices s and v . A shortest path $\rho(s, v)$ connecting a seed vertex s to an infectious or recovered vertex v is considered to be utilized if $d(s, v) = \pi(v)$, where $seed(v) = s$.

B. IM MODEL

Given a graph $G = (V, E)$, the IM problem deals with identifying a small set of seed vertices $S \subseteq V$ that result in the greatest spread of influence to other vertices in G , where $|S| = k \ll |V|$. This problem has two main components: (1) seed selection, which involves identifying the subset of seed vertices S ; and (2) influence diffusion, which describes the process by which influence is disseminated throughout the network over time. A number of algorithms have been proposed to solve the seed selection problem, including the following algorithms.

- The random model, which selects a set of k seed vertices uniformly at random.
- Centrality heuristics, which rank all vertices according to some centrality measure, then select the top k vertices as seed vertices. A commonly used centrality measure is degree centrality, which selects a set of k vertices with the highest degrees (i.e., the largest number of connections). Closeness centrality has been also used to select seed vertices. Closeness centrality is based on the assumption that vertices with shorter paths to other network vertices have a higher probability of spreading influence.

Each vertex is assigned one of two states: active and inactive. All vertices start as inactive except for the seed vertices. When the influence reaches a vertex, it becomes active. An active vertex u influences an inactive vertex v with probability p_{uv} . A number of algorithms have been proposed to model the diffusion process, including the following algorithms.

- *Linear Threshold Model*: Each vertex v has an assigned threshold θ_v chosen uniformly at random from $[0, 1]$. Every active neighbor vertex of an inactive vertex v contributes a certain weight, and if their sum exceeds the threshold value of vertex v , then v becomes active. That is, v is influenced by its neighbors when the fraction of its neighbors that are active is at least θ_v ($\sum_{u \in N(v)} p_{uv} \geq \theta_v$, where $N(v)$ is the set of neighbors of vertex v).
- *Independent Cascade (IC) Model*: At each time step, each active vertex u has a single chance of activating each of its inactive neighbor vertices with probability p .

For each network $G = (V, E)$ in our network datasets we identified a set of seed vertices S using three baseline seed selection algorithms: random, degree centrality, and closeness centrality. Then we ran the IC diffusion model several times for each network (one for each seed set). We defined two attributes for each vertex $v \in V$ as follows.

(1) $seed(v)$ denotes the seed vertex that influenced vertex v (directly or indirectly).

(2) $\pi(v)$ denotes the number of influenced vertices on the path from $seed(v)$ to v (i.e., the number of vertices that are influenced before in turn influencing vertex v).

At the end of each implementation, the total number of utilized shortest paths was computed as follows. Let $d(s, v)$ be the distance of a shortest path connecting vertices s and v . A shortest path $\rho(s, v)$ connecting a seed vertex s to an active vertex v is considered to be utilized if $d(s, v) = \pi(v)$.

VI. RESULTS

We applied various data dissemination scenarios to our network datasets to investigate shortest path utilization. During the SIR simulations, we defined an *infection path* as the path connecting a seed vertex s and an infectious (or recovered) vertex v . Similarly, we defined an *influence path* as the path connecting a seed vertex s and an active vertex v during IM simulations. We also took the following into consideration during our analyses.

- Shortest paths were identified by distance (not by unique paths). If the distance of an infection path or an influence path was equal to the shortest path between two vertices, then the shortest path was considered to be utilized. That is, given an infected (or influenced) vertex v and a seed vertex s such that $seed(v) = s$, a shortest path between s and v was considered to be utilized if $d(s, v) = \pi(v)$. This is important because more than one shortest path may exist between each pair of vertices in a given network.

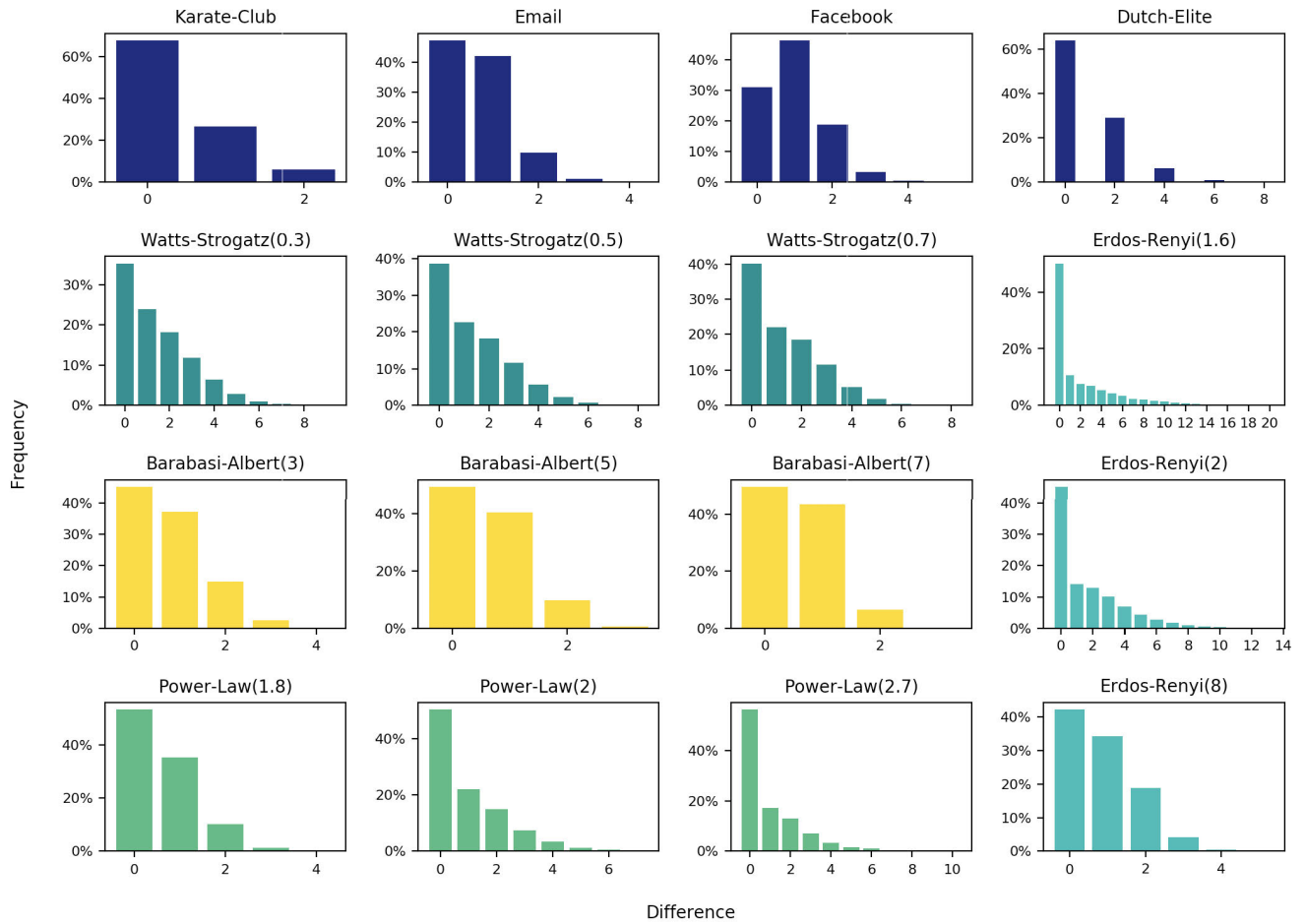


FIGURE 3. Distance differences between infection paths and shortest paths during SIR simulations.

- We focused our analyses on vertex pairs. That is, we only considered a single shortest path between each pair of vertices.

We also investigated the structural properties that could affect shortest path utilization. We focused on three distance properties of complex networks: diameter, average path length, and small-worldness (measured as the difference between a network size and its diameter).

A. SIR INFECTION-SPREADING MODEL

For each network in our network datasets, we replicated 100 SIR epidemic simulations (each starting from a randomly selected infected seed vertex), averaging the total fraction of infected vertices and infection paths throughout each simulation run. Each simulation run was 100 time steps long. During all simulations, the infection rate was 0.5 and the recovery rate was 0.1. These values were selected to guarantee transmission to all vertices within the specified simulation period. The average outbreak size (total number of infectious or recovered vertices) at the end of all simulations exceeded 90% in all networks.

Then, the lengths of the infection paths and the shortest paths were compared. The results are shown in Fig. 3.³ For each infectious or recovered vertex v , $\pi(v) - d(v, s) = 0$ if the infection path is equal to one of the shortest paths between the two vertices v and s (where s is the seed vertex that passed the infection to vertex v). Generally, $\pi(v) - d(v, s) = k$, where k is the difference between the length of the infection path and the shortest path.

Fig. 3 shows the percent of unutilized shortest paths in each network. For example, the SIR algorithm did not utilize 48% of the shortest paths in the Email network or about 70% of the shortest paths in the Facebook network. In the majority of networks, 50% or more of the shortest paths were not utilized (Facebook: Watts-Strogatz(0.3), Watts-Strogatz(0.5), Watts-Strogatz(0.7), Barabasi-Albert(3), Barabasi-Albert(5), Barabasi-Albert(7), Erdos-Renyi(1.6), Erdos-Renyi(2), Erdos-Renyi(8), Power-Law(2)). The exceptions were the Erdos-Renyi(1.6), Erdos-Renyi(2), and PowerLaw(2.7) networks. Notably, these three

³For comparison, a second implementation of this algorithm with a different infection rate value is included in Appendix A.

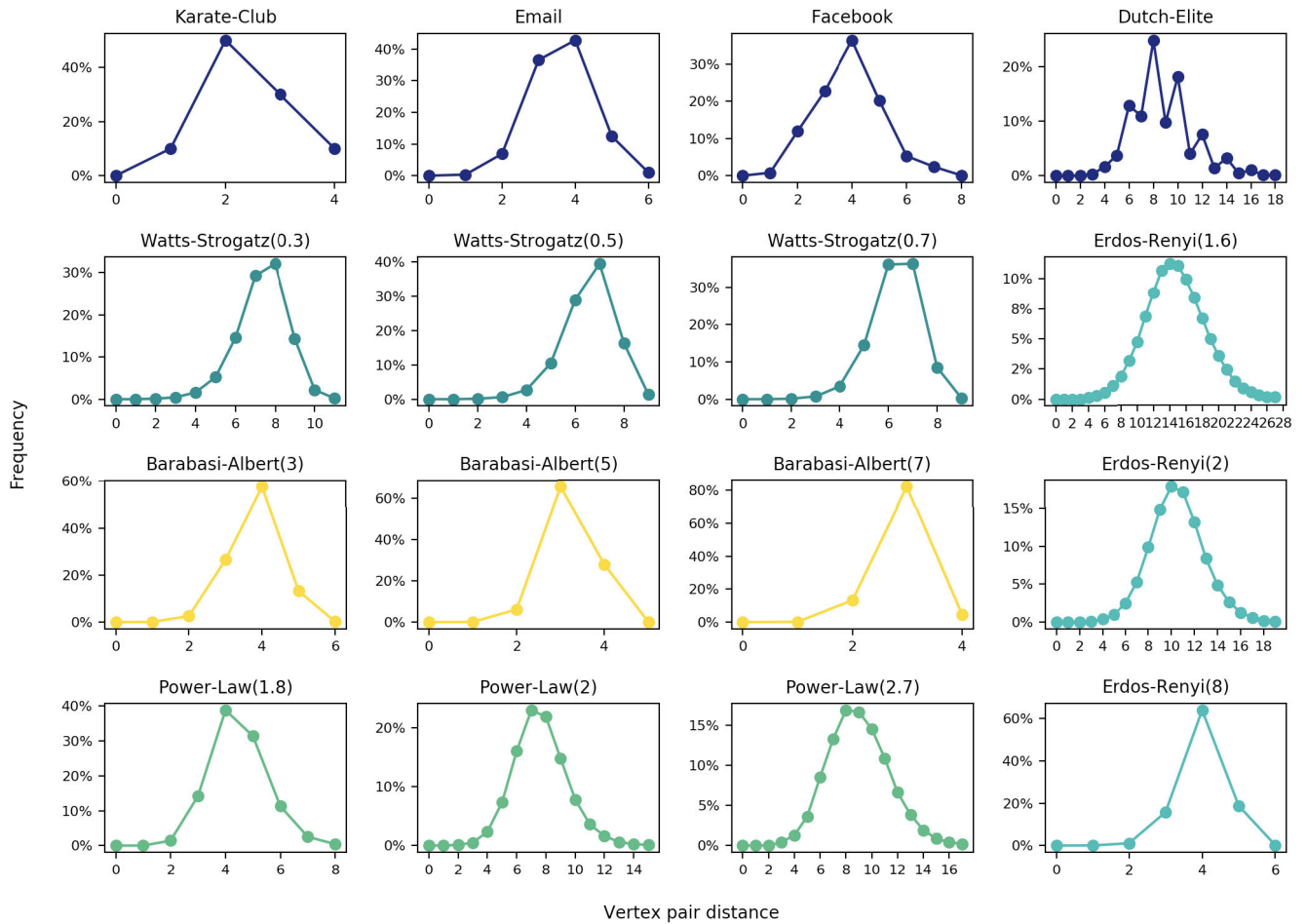


FIGURE 4. Lengths of unutilized shortest paths between vertex pairs in each network. In each panel, the x-axis ranges from zero to the network diameter (maximum distance).

networks had very large diameters compared with the network size (Table 2). When $diam(G) \geq size(G)$, where $size(G) = |V| + |E|$, this is an indication of the deviation of the network from the small-world structure.

Fig. 3 also shows the differences between the lengths of the infection paths and the shortest paths during each SIR simulation. In the Email network, 42% of the infection paths were one unit longer than the shortest paths, and 10% were two units longer than the shortest paths. The maximum difference between an infection path and a shortest path in the Email network was four. In the Watts–Strogatz(0.3) network, the percentages of infection paths with differences (compared with shortest paths) of one, two, three, four, five, and six were 24%, 18%, 12%, 7%, 3%, and 1%, respectively. The maximum difference between an infection path and a shortest path in the Watts–Strogatz(0.3) network was nine.

Fig. 4 shows the distances between vertex pairs for which the shortest path was not utilized. For example, in the Karate-Club network, the majority of unutilized shortest paths (about 50%) were of distance two. In most networks, shorter shortest paths were less utilized compared with longer shortest paths. For example, in all networks, all shortest paths with distances

equal to the network diameter were utilized. The results also show that the SIR algorithm was more likely to follow shortest paths when the distance between the vertex pair was too small.

Fig. 5 shows the eccentricities of the vertices (infectious or recovered) that exist at one end of an unutilized shortest path. In the Facebook network, 2% of infectious or recovered vertices that represented an end of an unutilized shortest path had eccentricity one, 63% had eccentricity two, 29% had eccentricity three, and 5% had eccentricity four. It is clear from the results shown in Fig. 5 that vertex eccentricities affect shortest path utilization. Specifically, the higher the eccentricity of a vertex v , the higher the probability of a shortest path that includes v as one of its endpoints not being utilized.

B. THE IM MODEL

For each network in our network datasets, we performed 100 IM simulations using one of the three baseline seed selection algorithms discussed above (random, degree centrality, and closeness centrality) and the IC influence diffusion model. The results of each simulation, including the fractions

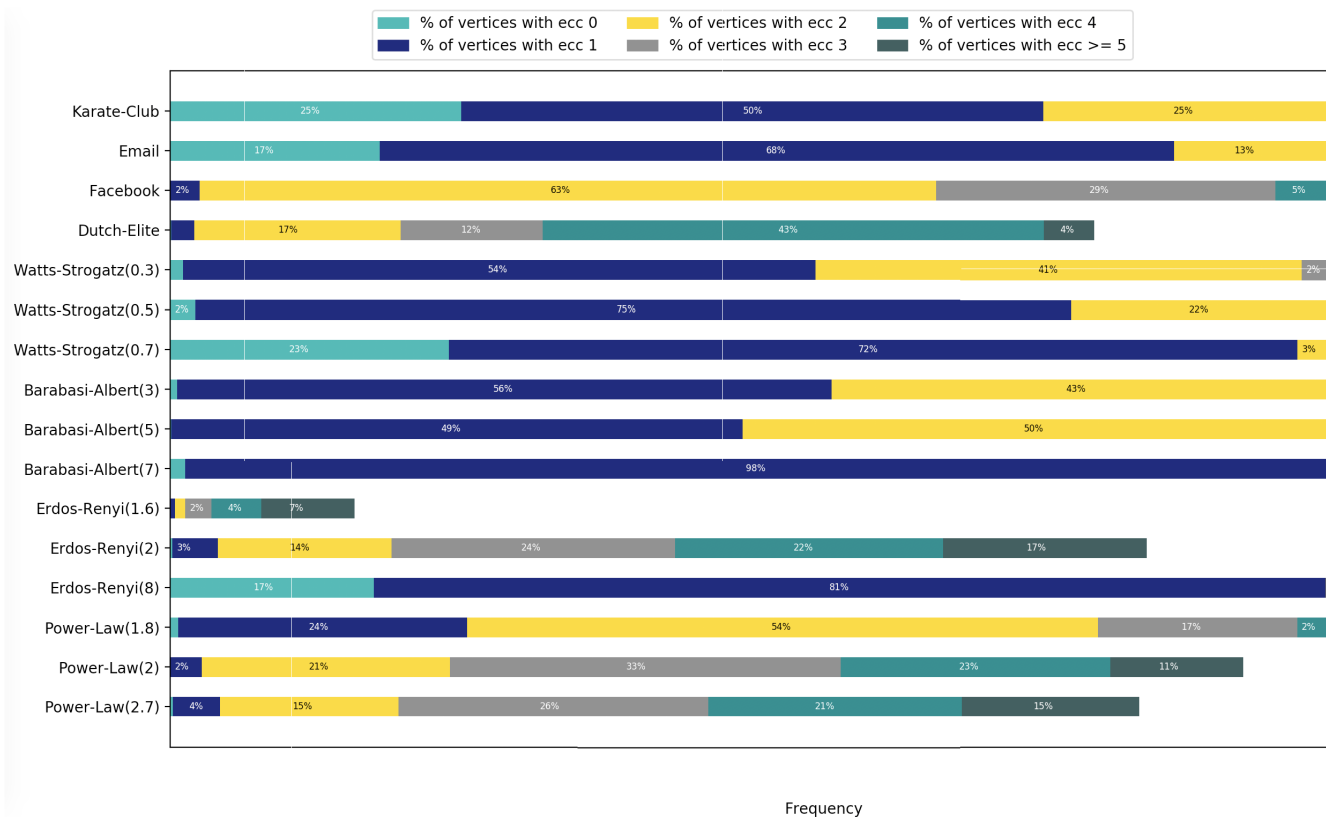


FIGURE 5. Eccentricities of vertices at endpoints of unutilized shortest paths in each network.

of influenced vertices and influence paths, were averaged throughout each simulation run. Each simulation run was 100 time steps long. During all simulations, the influence probability was 0.25 and the number of initial seed vertices represented 1% of the total number of vertices in the network. These values were selected to guarantee influence spread to the majority of other vertices within the specified simulation period. The results are summarized in Fig. 6.⁴

The first column in Fig. 6 shows the number of unutilized shortest paths in each network for each seed selection algorithm. Similar patterns could be observed to those seen in the case of the SIR algorithm. The main difference between the SIR and the IM algorithms is that, unlike the SIR algorithm, the IC algorithm gives each active vertex a single chance of activating each of its inactive neighboring vertices. As shown in Fig. 6, in some networks, 50% or more of the shortest paths were not utilized. These networks included the KarateClub, Email, Facebook, Barabasi–Albert(3), and Erdos–Renyi(8) networks. In other networks, 20% or more of the shortest paths were not utilized. These included the DutchElite, Watts–Strogatz(0.5), Watts–Strogatz(0.7), Barabasi–Albert(5), and Barabasi–Albert(7) networks. Between 1.2% and 18.4% of the shortest paths were utilized in the remaining networks

⁴For comparison, a second implementation of this algorithm with a different initial seed set size is included in Appendix B.

(Watts–Strogatz(0.3), Erdos–Renyi(2), PowerLaw(2), and PowerLaw(2.7)). All IM algorithms were able to utilize all the shortest paths in one single network (Erdos–Renyi(1.6)).

In addition, as shown in the Fig. 6, more shortest paths were utilized when the degree centrality or the closeness centrality was used for seed selection in the majority of the networks. On the other hand, random seed selection seemed to result in the highest shortest path unutilization. According to [8], influence spread is not very sensitive to the choice of algorithm when the value of p is large. However, the results in the second column in Fig. 6 suggest that the seed selection algorithm did have an impact on the outbreak size in some networks (KarateClub, DutchElite, Watts–Strogatz, Erdos–Renyi(1.6), Erdos–Renyi(2), PowerLaw(2), and PowerLaw(2.7)).

The third column in Fig. 6 shows the distribution of the differences between the influence paths and the shortest paths in each network (differences of one or more are shown in the figure). For example, using the random seed selection algorithm in the KarateClub network, 50% of influence paths were one unit longer than the shortest paths, 25% of influence paths were two units longer than the shortest paths, and 25% of influence paths were three units longer than the shortest paths. Using the degree centrality seed selection algorithm, 50%, 33%, and 16% of influence paths were one unit, two units, and three units longer than the shortest paths, respectively. The closeness centrality seed selection

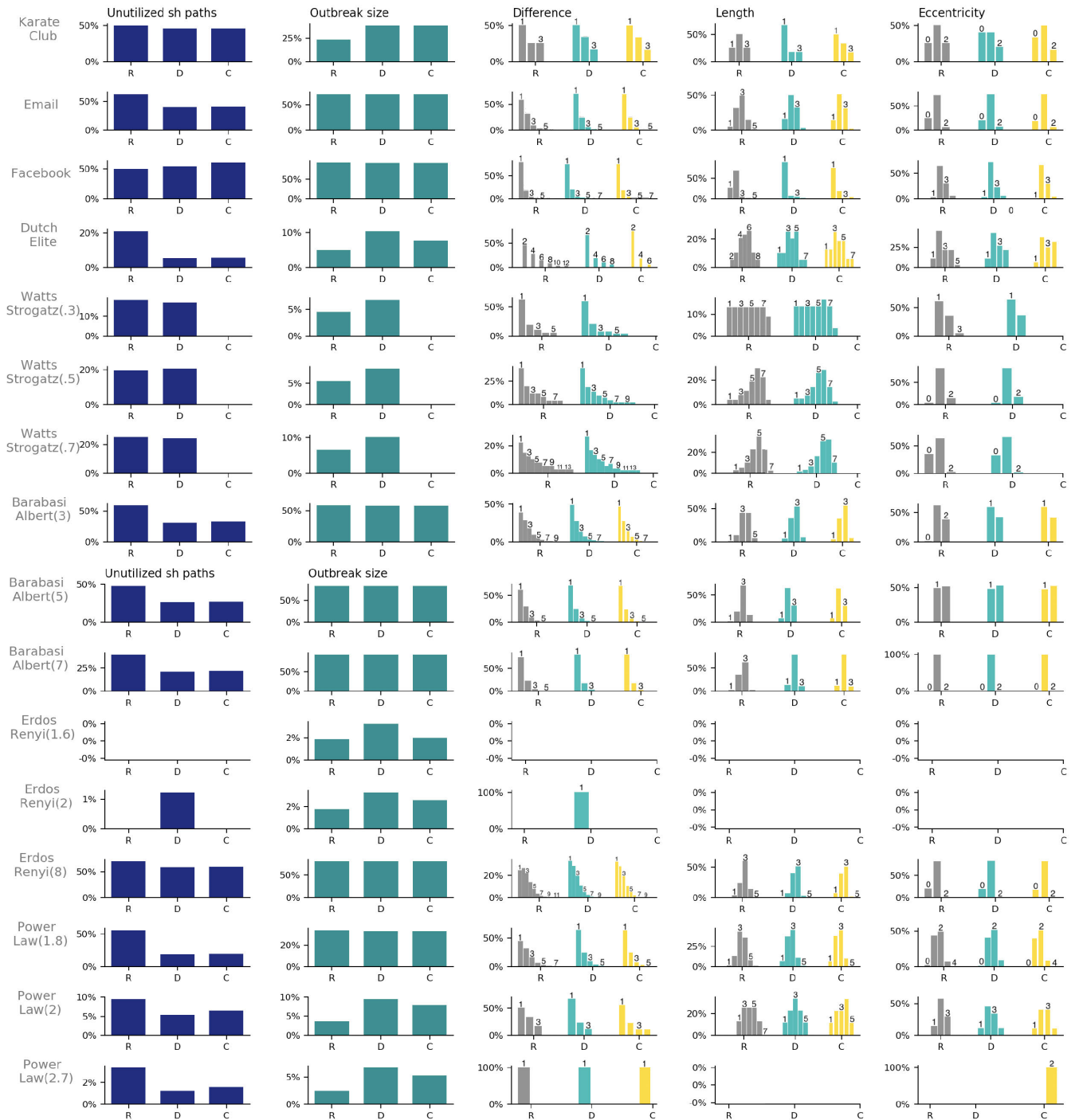


FIGURE 6. IM simulation results for each network using three seed selection algorithms: R (random seed selection algorithm), D (degree centrality seed selection algorithm), and C (closeness centrality seed selection algorithm).

algorithm achieved similar results to the degree centrality algorithm. Under the random and degree centrality seed selection algorithms, the maximum difference between an influence path and a shortest path was 14 (achieved by the Watts–Strogatz(0.7) network). With the closeness centrality seed selection algorithm, the maximum difference between an influence path and a shortest path was 10 (achieved by the Erdos–Renyi(8) network).

The fourth column in Fig. 6 shows the lengths of the unutilized shortest paths in each network. These results answer the following question: what is the shortest path distance between a vertex pair for which an influence path is longer than the shortest path? The lengths of most unutilized shortest paths were short (compared with the network diameter). For example, in the Facebook network, the majority of the unutilized shortest paths were of length one or two (68% using random

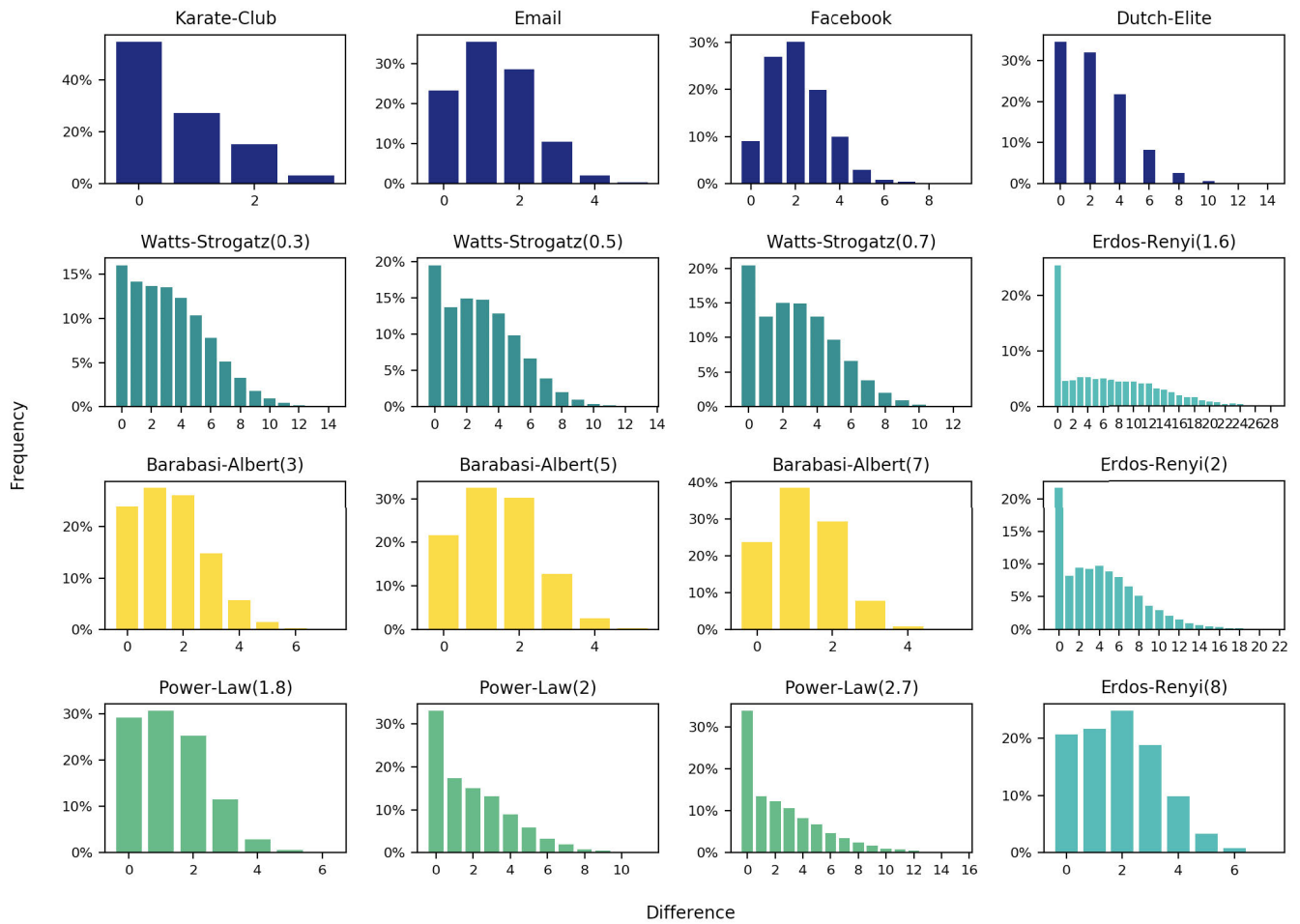


FIGURE 7. Distance differences between infection paths and shortest paths during SIR simulations.

seed selection, 89% using degree centrality seed selection, and 75% using closeness centrality seed selection).

The last column in Fig. 6 shows the eccentricity of vertices activated through an influence path that was not equal to the shortest path (i.e., vertices at an end point of an unutilized shortest path). In most networks, the majority of these vertices had medium eccentricities (with respect to the network radius). For example, in the Facebook network with the random seed selection algorithm, the majority of the vertices at an end point of an unutilized shortest path (63%) had eccentricities that were two units larger than the radius. Similarly, vertices with eccentricities that were one unit larger than the radius represented the majority of vertices existing at end points of unutilized shortest paths in the Facebook network with the degree and closeness centrality seed selection algorithms (70% and 66%, respectively). Moreover, as shown in Fig. 6, the seed selection algorithm did not affect the maximum eccentricity value in most networks (for example, the Email, Facebook, and Erdos–Renyi(8) networks).

C. THE ROLE OF NETWORK STRUCTURE

Table 3 shows the results of our investigation of the role of network structure in the utilization of shortest paths during

the selected data dissemination algorithms. We focused on a network’s distance properties (average path length, diameter, and small-worldness). Table 3 includes results for two implementations of each data dissemination algorithm. In the first and second SIR implementations, the infection rates were set to 0.5 and 0.25, respectively (both with 0.1 recovery rate). In the first and second IM implementations, the seed set sizes were set to 25% and 1% of the total number of vertices in each network, respectively (both with 0.25 influence rate). The IM results in Table 3 include the highest percentages of unutilized shortest paths achieved by any of the three algorithms (random, degree centrality, and closeness centrality) and their corresponding maximum difference values.

The networks in Table 3 were ordered according to their small-worldness (column 4), defined as the logarithmic difference between a network’s diameter and its size. The greater this difference, the stronger the small-world property of the network. According to their small-worldness, the network datasets were partitioned into two classes: small-world and non-small-world (a threshold of zero was selected for this partitioning).

We defined Δ_e as the maximum difference between an infection path and a shortest path in a given network.

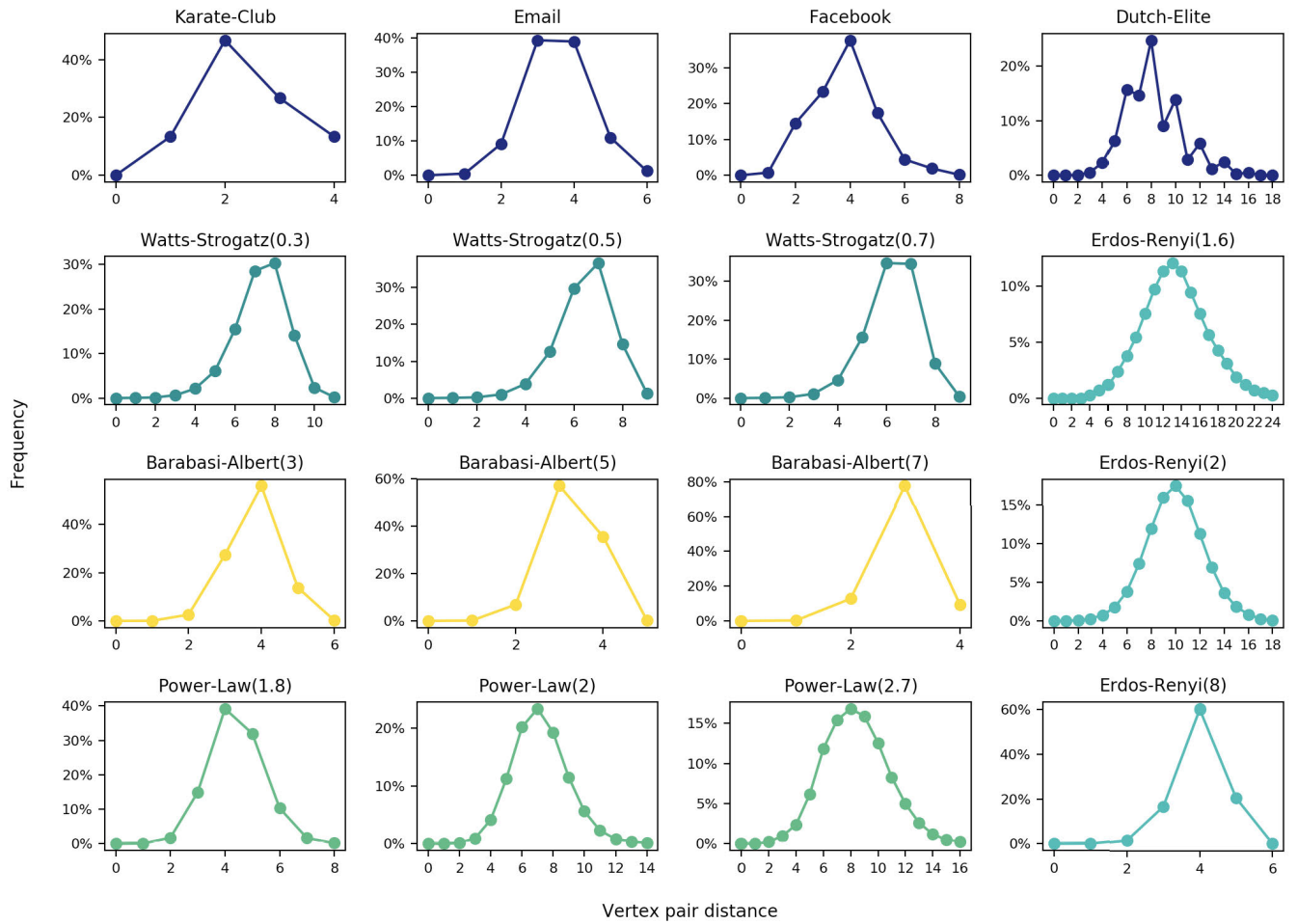


FIGURE 8. Lengths of unutilized shortest paths between vertex pairs in each network. In each panel, the x-axis ranges from zero to the network diameter (maximum distance).

TABLE 3. Utilization of shortest paths in the network datasets using two data dissemination models: the SIR (susceptible–infectious–recovered) model and the IM (influence maximization) model. Δ_e denotes the maximum difference between an infection path and a shortest path in a given network. Δ_l denotes the maximum difference between an influence path and a shortest path in a given network. Inf, influence; sh, shortest.

Network (G)	Avg. path length	$diam(G)$	$\log_2(\frac{size(G)}{diam(G)})$	SIR (Inf rate: 0.25)		SIR (Inf rate: 0.5)		IM (seed set: 25%)		IM (seed set: 1%)	
				Δ_e	unutilized sh paths	Δ_e	unutilized sh paths	Δ_l	unutilized sh paths	Δ_l	unutilized sh paths
Barabasi–Albert(7)	3	5	9.3	5	76%	3	50%	3	2%	5	39%
Barabasi–Albert(5)	3.3	5	8.9	5	78%	3	51%	3	3%	6	48%
Facebook	3.7	8	8.5	9	91%	5	69%	7	47%	7	60%
Barabasi–Albert(3)	3.8	6	7.3	7	76%	4	55%	4	3%	9	59%
Erdos–Renyi(8)	4	7	6.6	7	76%	5	53%	4	3%	11	70%
Email	3.6	8	4.7	5	73%	4	48%	3	8%	6	63%
Watts–Strogatz(0.7)	3.8	10	3.3	12	75%	8	53%	4	2%	14	25%
Watts–Strogatz(0.5)	6.1	11	1.9	13	80%	8	61%	4	3%	10	21%
KarateClub	2.4	5	1.8	3	40%	2	28%	3	18%	3	50%
PowerLaw(1.8)	4.5	11	1.6	6	71%	10	44%	3	2%	7	56%
Watts–Strogatz(0.3)	6.4	12	0.9	14	84%	9	64%	4	6%	5	18%
DutchElite	8.6	22	−9	14	65%	8	36%	2	0.3%	12	21%
PowerLaw(2.7)	7.4	21	−9.1	15	66%	7	50%	0	0%	1	3.3%
PowerLaw(2)	8.5	23	−11.7	11	67%	4	47%	2	0.5%	3	10%
Erdos–Renyi(2)	10.1	25	−12.9	21	78%	13	55%	0	0%	1	1.2%
Erdos–Renyi(1.6)	14.3	35	−23.3	28	75%	20	50%	0	0%	0	0%

Similarly, we defined Δ_l as the maximum difference between an influence path and a shortest path in a given network. Two observations can be made based on the results shown in Table 3. First, there was a positive correlation between

a network’s average path length and each of its Δ_e and Δ_l values. Second, the unutilization of shortest paths was lower among non-small-world networks in most implementations. Moreover, the difference between the length of an infection

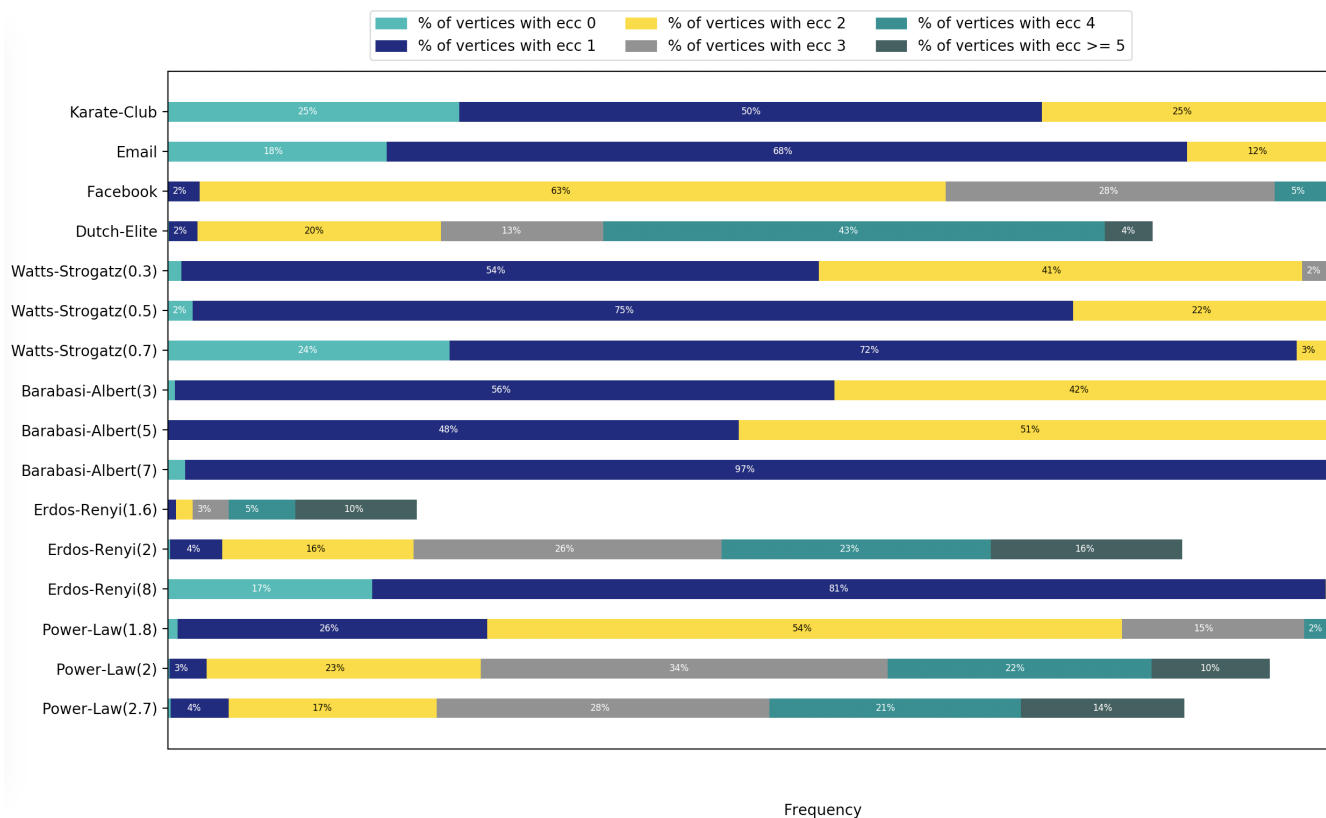


FIGURE 9. Eccentricities of vertices at endpoints of unused shortest paths in each network.

path and that of a shortest path was larger among non-small-world networks. The difference between the length of an influence path and that of a shortest path was smaller among non-small-world networks.

With the SIR model, the value of Δ_e seemed to correlate with the network diameter in all networks despite small-worldness (a network is small-world if the logarithmic difference between its size and its diameter is positive). The exceptions were the PowerLaw(2) and PowerLaw(2.7) networks. In addition, networks that did not exhibit the small-world property (DutchElite, Erdos–Renyi(1.6), Erdos–Renyi(2), PowerLaw(2), and PowerLaw(2.7)) achieved perfect (or almost perfect) shortest path utilization under one of the IM algorithms.

From a vertex perspective, our results show that vertex position and local properties play a part in shortest path utilization. Specifically, vertices with medium eccentricities seemed to be more likely to be at one end of an infection path (Fig. 5) or an influence path (last column in Fig. 6). Given a network G with radius $rad(G)$ and a diameter $diam(G)$, a vertex $v \in V(G)$ with an eccentricity $ecc(v)$ is considered to have medium eccentricity if $rad(G) < ecc(v) < diam(G)$. Vertex eccentricity is a measure of how close a vertex is to every other vertex in the network. Accordingly, our results indicate that a vertex v that is too close to or too far from every other vertex tends to be influenced through the expected path (the shortest path that connects v to the source vertex).

In our IM simulations, we explored how the position of the initial seed vertices influenced the whole network’s dynamics. In general, our results (see Fig. 6) show that a lack of shortest path utilization is not affected by the positions of the initial seed vertices. This suggests that the stochasticity of the model is a stronger determinant of a network’s dynamics than its structure.

VII. DISCUSSION

Many network analysis problems deal with identifying how information progresses from one vertex to another throughout a given network, for instance, disease spreading, influence maximization, facility location problems, routing protocol design, and relationship formation. A widely adopted assumption in most network analysis models is that data exchange processes trace the shortest paths in complex networks. This assumption has been the underlying concept of many network-analysis-based solutions. For example, the facility location problem, in which the goal is to determine the optimal location for a new facility (e.g., a hospital, a school, or a station) according to some defined criteria, depends on minimizing the distances traveled from each of the other vertices to the facility. Accordingly, centrality measures based on shortest paths are used to solve such problems.

Using a wide range of network structures, we investigated the utilization of shortest paths using several spreading processes. Our network datasets included four real-world

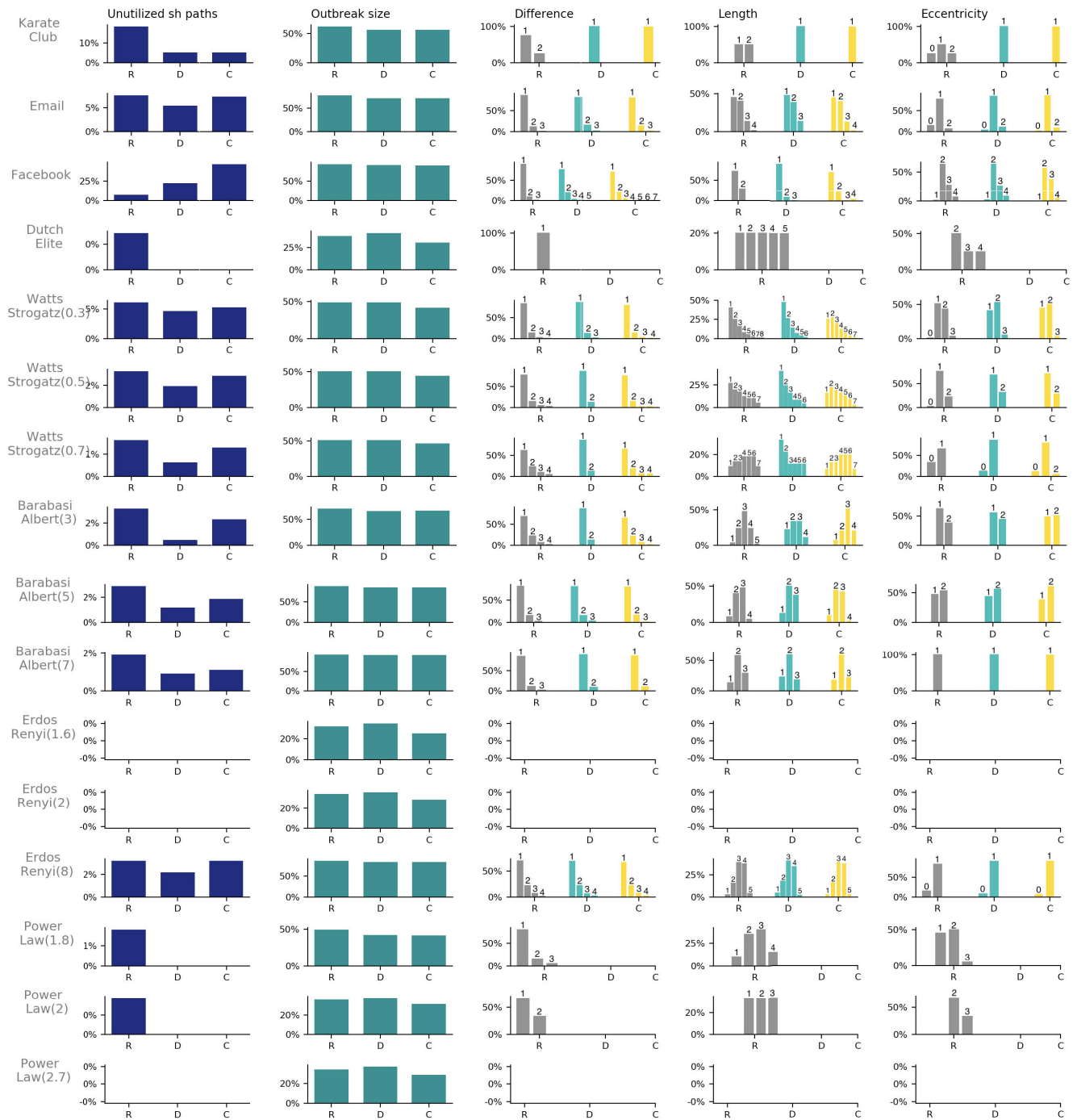


FIGURE 10. IM simulation results for each network using three seed selection algorithms: R (random seed selection algorithm), D (degree centrality seed selection algorithm), and C (closeness centrality seed selection algorithm).

social networks and four types of artificial networks, Watts–Strogatz, Barabasi–Albert, Erdos–Renyi, and PowerLaw. All networks of a certain type have some common features. However, they may exhibit slightly different topological properties. For example, real-world social networks are known to be small-world and scale-free; however, the DutchElite network did not seem to be small-world. Moreover, the topological properties of each artificial network are governed by its selected generation parameter or parameters. For example,

one of the PowerLaw networks was small-world, whereas the other two were not. Those topological properties seem to affect a network’s shortest path utilization. That is, the shortest path utilization of a network cannot be determined by the network type alone.

Our simulation results show that the utilization of shortest paths in complex networks may not be as common as assumed. This implies that longer paths can be as important (in some cases) as shortest paths. Our results show that at least

two factors clearly influence the shortest path utilization in a network: the network structure and the data dissemination algorithm. These results will affect the way we think about the role of vertices in data dissemination processes. For instance, the assignment of vertex importance may not need to depend on classic vertex-ranking measures such as degree, betweenness, closeness, or eccentricity centralities.

However, the findings of this work were subject to several limitations. First, the experiments were limited to a specific set of network types. More network types, as well as more variations of each network type, would need to be included to enhance the accuracy of our findings. Moreover, in our analyses, we focused on two data dissemination algorithms. It would be interesting to include other types of data dissemination algorithms to obtain a better overview of how networks behave. The analyses conducted in this work represent only the beginning of research in this direction. Further analyses should investigate more network structures and vertex properties.

VIII. CONCLUSION

The study of networks enables understanding of their statistics, structures, and dynamics. A network's structure has been found to highly influence its dynamics, including data dissemination between vertices, pathogen spread throughout a network, and relationship formation. A very widely used assumption in most network analysis models is that traffic follows the shortest paths connecting pairs of nonneighboring vertices. Accordingly, network analysis models are often based on the idea that information is passed from one vertex to another only along these shortest paths, for example, graph centrality measures, community extraction algorithms, and core-periphery detection algorithms. However, this perspective is very restricted and can be misleading as a consequence of its focus on shortest path communications. In this work, we investigated shortest path utilization in real and artificial networks using a set of simulations. Although our analysis methodology was general in nature, we focused on simulating two data dissemination algorithms: the SIR model and the IM model.

We conclude that the shortest path utilization in a network is influenced by a combination of factors that are related to its structure and the data dissemination algorithm. We also found that network type was not a good indicator of shortest path utilization. Our results have significant implications for network analysis, network modeling, and network generation.

Future research in this direction should investigate more data dissemination models and scenarios. Future analysis should also include more network types and examine more network structural properties, for example, the clustering coefficient of a given network.

APPENDIX A

In the SIR implementation, the infection rate was set to 0.25 and the recovery rate was set to 0.1. See Figs. 7, 8, and 9 for the results.

APPENDIX B

In the IM implementation, the seed set size was set to 25% of the total number of vertices in each network. The influence rate used was 0.25. See Fig. 10 for the results.

REFERENCES

- [1] M. Granovetter, "The strength of weak ties," *Amer. J. Sociol.*, vol. 78, no. 6, pp. 1360–1380, May 1973.
- [2] D. J. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.
- [3] J. Kleinberg, "The small-world phenomenon: An algorithmic perspective," in *Proc. 32nd Annu. ACM Symp. Theory Comput.*, May 2000, pp. 163–170.
- [4] L. Lu, "The diameter of random massive graphs," in *Proc. 12th Annu. ACM-SIAM Symp. Discrete Algorithms*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, Jan. 2001, pp. 912–921.
- [5] F. Chung and L. Lu, "The average distances in random graphs with given expected degrees," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 25, pp. 15879–15882, Dec. 2002.
- [6] M. E. J. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, 2003.
- [7] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, "Group formation in large social networks: Membership, growth, and evolution," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2006, pp. 44–54.
- [8] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jun. 2009, pp. 199–208.
- [9] W. Nawaz, K. U. Khan, and Y.-K. Lee, "Shortest path analysis in social graphs," 2014, *arXiv:1412.7760*. [Online]. Available: <http://arxiv.org/abs/1412.7760>
- [10] G. Strona, C. J. Carstens, P. S. A. Beck, and B. A. Han, "The intrinsic vulnerability of networks to epidemics," *Ecol. Modell.*, vol. 383, pp. 91–97, Sep. 2018.
- [11] S. Sottile, O. Kahramanoğlu, and M. Sensi, "How network properties and epidemic parameters influence stochastic SIR dynamics on scale-free random networks," 2020, *arXiv:2011.10595*. [Online]. Available: <http://arxiv.org/abs/2011.10595>
- [12] S. Zhu and D. Levinson, "Do people use the shortest path? An empirical test of Wardrop's first principle," *PLoS ONE*, vol. 10, no. 8, Aug. 2015, Art. no. e0134322.
- [13] L. C. Freeman, "Centrality in social networks conceptual clarification," *Soc. Netw.*, vol. 1, no. 3, pp. 215–239, Jan. 1978.
- [14] S. P. Borgatti, "Centrality and network flow," *Soc. Netw.*, vol. 27, no. 1, pp. 55–71, Jan. 2005.
- [15] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, Jun. 2006.
- [16] J. W. Pinney and D. R. Westhead, "Betweenness-based decomposition methods for social and biological networks," *Interdiscipl. Stat. Bioinf.*, vol. 25, pp. 87–90, Jul. 2006.
- [17] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech., Theory Exp.*, vol. 2008, no. 10, Oct. 2008, Art. no. P10008.
- [18] P. Holme, "Core-periphery organization of complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 72, no. 4, Oct. 2005, Art. no. 046111.
- [19] S. P. Borgatti and M. G. Everett, "Models of core/periphery structures," *Social Netw.*, vol. 21, no. 4, pp. 375–395, Oct. 2000.
- [20] M. P. Rombach, M. A. Porter, J. H. Fowler, and P. J. Mucha, "Core-periphery structure in networks," *SIAM J. Appl. Math.*, vol. 74, no. 1, pp. 167–190, 2014.
- [21] K. Stephenson and M. Zelen, "Rethinking centrality: Methods and examples," *Soc. Netw.*, vol. 11, no. 1, pp. 1–37, Mar. 1989.
- [22] L. C. Freeman, S. P. Borgatti, and D. R. White, "Centrality in valued graphs: A measure of betweenness based on network flow," *Social Netw.*, vol. 13, no. 2, pp. 141–154, Jun. 1991.
- [23] J. Kleinfield, "Could it be a big world after all? The six degrees of separation myth," *Society*, vol. 12, pp. 5–2, Apr. 2002.
- [24] S. Schnettler, "A small world on feet of clay? A comparison of empirical small-world studies against best-practice criteria," *Social Netw.*, vol. 31, no. 3, pp. 179–189, Jul. 2009.

- [25] S. Schnettler, "A structured overview of 50 years of small-world research," *Social Netw.*, vol. 31, no. 3, pp. 165–178, Jul. 2009.
- [26] S. Milgram, "The small world problem," *Psychol. Today*, vol. 2, no. 1, pp. 60–67, May 1967.
- [27] M. E. Newman, "The structure of scientific collaboration networks," in *The Structure and Dynamics of Networks*. Princeton, NJ, USA: Princeton Univ. Press, Jan. 2011, pp. 221–226.
- [28] H. Alrasheed and F. F. Dragan, "Core-periphery models for graphs based on their δ -hyperbolicity: An example using biological networks," *J. Algorithms Comput. Technol.*, vol. 11, no. 1, pp. 40–57, Mar. 2017.
- [29] C. Wilson, B. Boe, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao, "User interactions in social networks and their implications," in *Proc. 4th ACM Eur. Conf. Comput. Syst.*, Apr. 2009, pp. 205–218.
- [30] J.-P. Onnela and N. A. Christakis, "Spreading paths in partially observed social networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 85, no. 3, Mar. 2012, Art. no. 036106.
- [31] X. Zuo, J. Blackburn, N. Kourtellis, J. Skvoretz, and A. Iamnitchi, "The influence of indirect ties on social network dynamics," in *Proc. Int. Conf. Social Informat.* Cham, Switzerland: Springer, Nov. 2014, pp. 50–65.
- [32] N. A. Christakis and J. H. Fowler, "The spread of obesity in a large social network over 32 years," *New England J. Med.*, vol. 357, no. 4, pp. 370–379, 2007.
- [33] J. H. Fowler and N. A. Christakis, "Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham heart study," *BMJ*, vol. 337, pp. 1–15, Dec. 2008.
- [34] F. M. Shatu, "Built environment impact on pedestrian route choice behaviour: Shortest vs. least directional change routes," Ph.D. dissertation, Queensland Univ. Technol., Brisbane City, QLD, Australia, 2019.
- [35] W. Tang and D. M. Levinson, "Deviation between actual and shortest travel time paths for commuters," *J. Transp. Eng. A Syst.*, vol. 144, no. 8, Aug. 2018, Art. no. 0401804.
- [36] A. Gulyás, Z. Heszberger, and J. Biró, "Everybody loves roundabouts," in *Paths*. Cham, Switzerland: Birkhäuser, 2021, pp. 5–10.
- [37] A. Gulyás, Z. Heszberger, and J. Biró, "The universal nature of paths," in *Paths*. Cham, Switzerland: Birkhäuser, 2021, pp. 45–65.
- [38] T. Simas, R. B. Correia, and L. M. Rocha, "The distance backbone of complex networks," 2021, *arXiv:2103.04668*. [Online]. Available: <http://arxiv.org/abs/2103.04668>
- [39] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, p. 47, Jan. 2000.
- [40] J. Abello, A. L. Buchsbaum, and J. R. Westbrook, "A functional approach to external graph algorithms," in *Proc. Eur. Symp. Algorithms*. Berlin, Germany: Springer, Aug. 1998, pp. 332–343.
- [41] *Network Repository*. Accessed: May 1, 2021. [Online]. Available: <http://networkrepository.com/soc-karate.php>
- [42] *Alex Arenas*. Accessed: Mar. 1, 2021. [Online]. Available: <http://deim.urv.cat/aarenas/data/welcome.htm>
- [43] V. Batagelj and A. Mrvar. *Pajek Datasets*. Accessed: May 1, 2021. [Online]. Available: <http://vlado.fmf.uni-lj.si/pub/networks/data/>
- [44] *Stanford Large Network Dataset Collection*. Accessed: May 1, 2021. [Online]. Available: <https://snap.stanford.edu/data/>
- [45] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.
- [46] W. Aiello, F. Chung, and L. Lu, "A random graph model for massive graphs," in *Proc. 32nd Annu. ACM Symp. Theory Comput. (STOC)*, 2000, pp. 171–180.
- [47] A. Brady, "A compact routing scheme for power-law networks using empirical discoveries in power-law graph topology," Ph.D. dissertation, Tufts Univ., Medford, MA, USA, 2005.
- [48] R. M. Anderson, B. Anderson, and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*. London, U.K.: Oxford Univ. Press, 1992.

HEND ALRASHEED received the B.S. and M.S. degrees in computer science from King Saud University, Riyadh, Saudi Arabia, in 2004 and 2009, respectively, and the Ph.D. degree in computer science from Kent State University, Kent, OH, USA, in 2018. She is currently an Assistant Professor with the College of Computer and Information Sciences, King Saud University. Her research focuses on analyzing structural properties of complex networks. Problems she investigates lie at the intersection of graph theory and network science. Her research interest includes software engineering measurements. Specifically, problems related to modeling software programs as flow graphs and analyzing their properties.

• • •