Local network connectivity optimization: An evaluation of heuristics applied to complex spatial networks, a transportation case study, and a spatial social network

Jeremy Auerbach[1] **and Hyun Kim**[2]

[1]**Colorado State University, Department of Environmental and Radiological Health Sciences, Fort Collins, CO 80523, USA**

[2]**University of Tennessee, Department of Geography, Knoxville, TN 37996, USA**

Corresponding author:

Jeremy Auerbach[1]

Email address: jeremy.auerbach@colostate.edu

ABSTRACT

Optimizing global connectivity in spatial networks, either through rewiring or adding edges, can increase the flow of information and increase the resilience of the network to failures. Yet, rewiring is not feasible for systems with fixed edges and optimizing global connectivity may not result in optimal local connectivity in systems where that is wanted. We describe the local network connectivity optimization problem, where costly edges are added to a systems with an established and fixed edge network to increase connectivity to a specific location, such as in transportation and telecommunication systems. Solutions to this problem maximize the number of nodes within a given distance to a focal node in the network while they minimize the number and length of additional connections. We compare several heuristics applied to random networks, including two novel planar random networks that are useful for spatial network simulation research, a real-world transportation case study, and a set of real-world social network data. Across network types, significant variation between nodal characteristics and the optimal connections was observed. The characteristics along with the computational costs of the search for optimal solutions highlights the need of prescribing effective heuristics. We offer a novel formulation of the genetic algorithm, which outperforms existing techniques. We describe how this heuristic can be applied to other combinatorial and dynamic problems.

INTRODUCTION

Spatial networks have become more popular as the interest in networks has spread into more fields and spatial data, and the computational power and methods to analyze it, have become more accessible. In terms of analysis, spatial network optimization has been at the forefront and focused on increasing network connectivity and information flow (Schrijver, 2002; Wu et al., 2004). Heuristics have been developed to rearrange existing networks or creating new ones that optimize the topology of the network for synchronizability (Khafa and Jalili, 2019). Several effective methods have also been developed to add new edges to a network that minimize the average shortest path distance (Meyerson and Tagiku, 2009), minimize the network diameter (Demaine and Zadimoghaddam, 2010), or maximize the network's centrality (Jiang et al., 2011) or connectivity (Alenazi et al., 2014).

While this optimization of spatial networks' global characteristics , optimizing local existing network connectivity around a specific node or location with the introduction of costly new edges has not been explored and yet is important in several domains. For example, increasing an existing network's connectivity around a focal node while minimizing the costs associated with the number and lengths of additional connections is essential in network layout planning for telecommunications and computer systems (Resende and Pardalos, 2006; Donoso and Fabregat, 2007) and increasing or slowing the spread of information or diseases in social networks (Gavrilets et al., 2016; Eubank et al., 2004). This local network

45 connectivity problem is particularly important with transportation planning in urban environments, where
46 the weights of the network edges can be physical distances or riderships and future street connections
47 or transportation lines can be significantly costly and impact flow to established facilities. For example,
48 planners can optimize thoroughfare connectivity around schools to foster student walking and biking
49 while reducing busing costs (Auerbach et al., 2021) and increase accessibility and patient travel time to
50 health care facilities (Branas et al., 2005).

51   The search for new edges that maximize connectivity to a focal node and minimize the costs of these
52 new edges is not well understood and this search for optimal solutions can become costly when networks
53 are large and complex. To fill this knowledge gap, we compare a set of heuristics to optimize local
54 network connectivity applied to real-world networks and randomly generated ones. These heuristics are
55 drawn from Mladenović et al. (2007) review of combinatorial heuristics and from location models that
56 include a spatial component (Brimberg and Hodgson, 2011). We also offer a genetic algorithm with a
57 novel chromosome formulation where the genes are not properties of a specific variable but weights for
58 the probability to move in a given dimension across the solution space.

59   These optimization heuristics are then applied to randomly generated networks that vary in complexity
60 and size to evaluate their efficacy in finding the optimal new connections that maximize local connectivity.
61 Included in this set of random graphs we provide two novel formulations of random planar networks based
62 on the Voronoi diagram and the Delaunay triangulation. To complement the random network analysis, the
63 network connectivity optimization methods are also applied to two real-world case studies, one from urban
64 transportation planning and another from social network analysis. In this study, we show that optimization
65 heuristics are preferred for the analysis and practice due to the nonlinearity of the solution space and the
66 optimal solution's dependence on nodal characteristics, such as distance to the focal node. The novel
67 genetic algorithm outperformed the other heuristics as it was able to move from suboptimal solutions
68 and explore distance solutions quicker. This is important as researchers and engineers are working with
69 networks or growing complexity and size.

70   The organization of this paper is as follows. The next section describes the formulation of the
71 connectivity problem in more detail, the local search methodology, and the optimization heuristics (see
72 Appendix A for the specific pseudocode of the optimization algorithms). This is followed by a section
73 that details the data used for the study including descriptions of the random networks, the transportation
74 case study street networks, and the social network data. Results of the heuristics applied to the random
75 networks and the case studies are then presented. The paper concludes with a detailed discussion of these
76 heuristic results, the further implications of these techniques for urban transportation planning, and future
77 work for this avenue of research.

METHODS AND DATA

Formulation of the Network Connectivity Problem

For the description of the optimization methodology the following nomenclature will be used (see Table 1).
In connectivity optimization, network nodes are first segmented and assigned to 'close' and 'distant' sets
by a chosen threshold distance D from the network's focal node F. The number of nodes v is a network is
N, and nodes are separated into two sets based on their shortest network path distances to the focal node,
$d(v,F)$. The nodes that are within this distance are assigned to the 'close' set, $N_C \subset N$, i.e., $v \in N_C$ if
$d(v,F) \leq D$ and $F \in N_C$ (see Figure 1 (A)). The nodes that are outside the threshold shortest network
path distance to the focal node, D, are assigned to the 'distant' set, $N_D \subset N$, i.e., $v \in N_D$ if $d(v,F) > D$.
When a new connection is added to the network, the shortest path distance from each distant node to the
focal node is recalculated. If there are any distant nodes that are now within the threshold distance to the
focal node they are assigned to the new set $N_{i,j}^C$. For example, if a new connection is established between
distant node i and close node j, then $k \in N_{i,j}^C$ if $k \in N_D$ and $d(k,i)+d(i,j)+d(j,F) \leq D$ (Figure 1 (C)).
The benefit of this new connection is $B(i,j)$ and the cost associated with the new connection is $C(i,j)$.
The optimal solution is the solution with the greatest benefit, or number of new nodes now within the

distance to the focal node which can be expressed as the bi-objective function

$$O^* = \max(B(i,j)) + \min(C(i,j))$$
$$\text{s.t. } \sum_i v_i = 1, i \in N_C$$
$$\sum_j v_j = 1, j \in N_D \tag{1}$$

where benefits dominate costs. For example, if $B(i,j) = B(m,n)$ and $C(i,j) < C(m,n)$, then the optimal additional edge is between i and j. For nondominated solutions, we select the solution that minimizes the distance to the so-called ideal point. The ideal point represents the solution that simultaneously maximizes the benefit and minimizes the cost. For the analysis in this paper the formulation of the objective function is as follows. For a new edge between i and j the number of nodes in $N_{i,j}^C$ set is the benefit of this new connection, $B(i,j) = |N_{i,j}^C|$, and the cost of the new connection is the length of the edge $C(i,j) = d(i,j)$. Therefore,

$$O^* = \max(|N_{i,j}^C|) + \min(d(i,j))$$
$$\text{s.t. } \sum_i v_i = 1, i \in N_C$$
$$\sum_j v_j = 1, j \in N_D \tag{2}$$

Most of the heuristics presented below are dependent on the number of iterations (t) and terminate when the solutions converge, $O^t = O^{t-1}$, or the solution does not improve, $O^t < O^{t-1}$.

Local Search Methodology

For a network of size N the number of new connections to evaluate has an upper bound of $N^2/4$, therefore heuristics may be employed to identify (nearly) optimal solutions quicker than an exhaustive search as networks get larger. These optimization algorithms require a search space to explore and using nodal characteristics we create such a multidimensional solution space (see Table 2). These nodal characteristics are explored to find the critical network properties for connectivity optimization and their impact on the performance in finding the optimal solution. As several nodes in a network can potentially have the same nodal characteristic values the local searches include a random shuffling routine. In detail, the following individual node characteristics were used to create the dimensions of the solution space: (i) distance to the focal node, (ii) degree centrality, (iii) closeness centrality, (iv) betweenness centrality, (v) eigenvector centrality, (vi) pagerank centrality, (vii) weighted clustering coefficient, and (viii) the neighbor nodes directly connected to the node. Nodes were sorted by their distance to the focal node and moving in this solution dimension may result in lower connectivity length costs but may not maximize the number of nodes ultimately connected to the focal node. In contrast, sorting nodes by their centrality, i.e., the importance of the node, could result in maximizing the number of nodes within the specified distance to the focal node but with the possibility of higher connectivity length costs compared to selecting nodes by other characteristics. To avoid these two extreme possibilities, several commonly used measures of centrality are explored: degree, the number of edges incident to a node; closeness centrality, the average length of the shortest path between the node and all other nodes in the network (Bavelas, 1950); betweenness centrality, the frequency of a node included in the shortest paths between all other node pairs (Freeman, 1977); eigenvector centrality, which is a relative sorting of nodes such that nodes with high values are connected to other nodes with high values (Newman, 2008); and pagerank centrality, a variant of eigenvector centrality that sorts nodes based on their probability of being connected to a randomly selected node and which is commonly used in web-page rankings (Brin and Page, 1998). The weighted clustering coefficient of a node is the count of the triplets in the neighborhood of the node and accounts for the weights of the edges times the maximum possible number of triplets that could occur (Barrat et al., 2004). The weights used here are the spatial distances between nodes $w_{ij} = d(i,j)$. The nodes that were directly connected to the node under evaluation were also used.

Network Connectivity Optimization Heuristics

The following techniques were implemented for the network connectivity optimization study from their extensive use in optimization: hill climbing with random restart (Russell and Norvig, 2004); stochastic
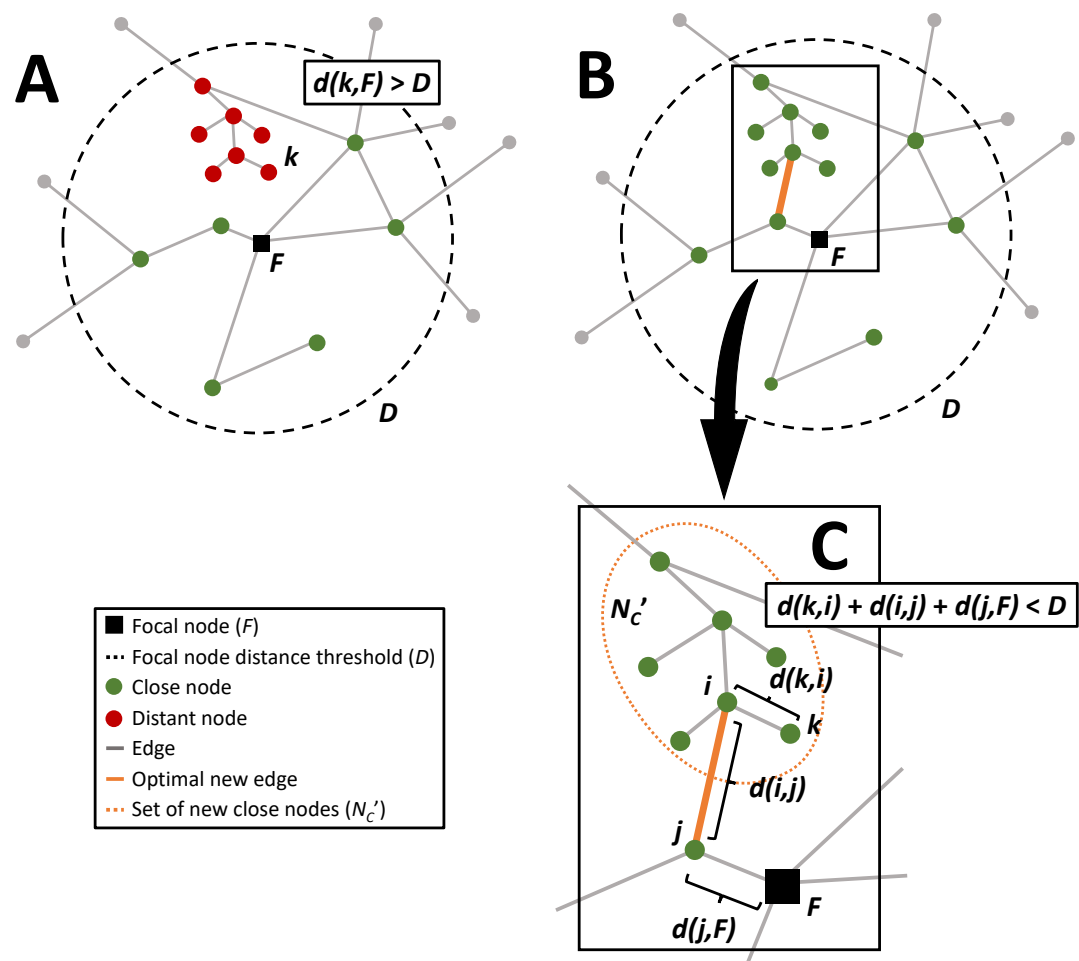
Figure 1. Diagram of the sequence of the network connectivity optimization problem. The close nodes that are within a threshold network distance (orange dashed circle) from the focal node (black square) are colored green, distant nodes that could be within the threshold network distance with additional edges are colored red, and the gray distant nodes are outside the threshold distance regardless of any additional connections. Figure (A) is an example graph, (B) shows the same graph with the optimal new connection that maximizes the number of additional nodes within the threshold network distance and minimizes the length of the new connection, and the inset (C) highlights this optimal connection, between nodes i and j.

113   hill climbing (Greiner, 1992); hill climbing with a variable neighborhood search (Mladenović and Hansen,
114   1997); simulated annealing, which has a history of applications in graph problems (Kirkpatrick et al.,
115   1983; Johnson et al., 1989; Kirkpatrick, 1984); and genetic algorithms, which has been successfully used
116   for combinatorial optimization (Anderson and Ferris, 1994; Jaramillo et al., 2002). A Tabu heuristic was
117   not employed as it has been observed to be an inferior method for multi-objective optimization problems
118   compared to simulated annealing and genetic algorithms (Golden and Skiscim, 1986; Kim et al., 2016).
119   Parameter selection was simplified for easy comparison of the methods (see the Supplemental Information
120   for the algorithms). To ensure that the heuristics did not converge on suboptimal solutions due to the
121   initial starting values, random restart, i.e., randomly selecting initial nodes to avoid local optima and
122   running the routine until the optimal solution is found, was used.
123   **Exhaustive search (ES).** The exhaustive search calculates the solution for every pair of distant and close
124   nodes for a network (see Algorithm 1 in the Supplemental Information). While this approach ensures
125   that the optimal edges are found, as the number of nodes increases and therefore the number of possible
126   connections between close and distant nodes increases, it can become computationally expensive and
127   timely to implement. Since the results of ES provide the optimal solution, the times it takes to find

128 evaluate all solutions are used to benchmark the other heuristics.

Hill climbing (HC). The solution space was observed to be hilly from the exhaustive search results, so several modifications were introduced to the hill climbing technique to avoid getting stuck in suboptimal solutions (Algorithm 2 in the Supplemental Information). A stochastic hill climbing **(HCS)**, an advanced search method based on HC, routine is also explored where the selection of nodes for the next iteration is randomly picked with

$$\text{probability}(i,j) = \frac{O(i,j)}{\sum_{(m,n)} O(m,n)}, \tag{3}$$

which terminates when an improved solution is no longer found (Algorithm 3 in the Supplemental Information). A hill climbing algorithm is coupled with a variable neighborhood **(HCVN)** where the size of the neighborhood starts with the nearest neighbors ($\eta = 1$) and is updated as follows:

$$\eta = \begin{cases} 1 & \text{if } O^t > O^{t-1} \\ \eta + 1 & \text{if } O^t \leq O^{t-1} \end{cases}, \tag{4}$$

129 and the HCVN method terminates after n_{max} is reached (Algorithm 4 in the Supplemental Information).

Simulated annealing (SA). As a meta-heuristic approach, the simulated annealing method randomly selects an initial solution from the solution space to avoid entrapment in a local optima. At each iteration, the heuristic evaluates the neighboring solutions and if it does not find an improved solution, it moves to a new solution with the following probability:

$$\text{probability}(i,j) = \exp\left(-\frac{O^{t-1} - O(i,j)}{t}\right). \tag{5}$$

130 The distance of the move decreases with the number of iterations until a better solution is no longer found
131 (Algorithm 5 in the Supplemental Information).

Genetic algorithm (GA). The genetic algorithm begins with a population of P randomly selected solutions with a set of chromosomes composed of genes which represent the weights of selecting a neighbor and are all initialized to unity (Algorithm 6 in the Supplemental Information). During each iteration of the method, solution scores (fitnesses) are computed by

$$f(i,j) = \frac{O(i,j)}{\sum_{(m,n)} O(m,n)}, \tag{6}$$

and a new generation of solutions are selected based on the following probability condition

$$\text{probability}(i,j) = \frac{s * f(i,j) + (1-s)}{\sum_{(m,n)} (s * f(m,n) + (1-s))}, \tag{7}$$

where s is the selection coefficient. Weak selection, $s \ll 1$, is used to ensure that random mutations impact solution frequency. Crossover is conducted by alternating the weights for the offspring from each parent, also known as cycle crossover (Oliver et al., 1987). Mutations are introduced at a low rate $\mu \ll 1$ for each gene and increase the nodal characteristic selection weight by one. The probability that characteristic m is used to find a neighbor for node i is given by

$$\text{probability}(\text{characteristic}) = \frac{\text{gene}(i,m)}{\sum_k \text{gene}(i,k)/K}, \tag{8}$$

132 where K is the total number of nodal characteristics. This formulation ensures that the nodal characteristics
133 that improve the solution increase in weight which results in a greater probability they will be selected
134 for neighborhood exploration, and ultimately reduces the size of the neighborhood search. Among these
135 methods, the genetic algorithm presented here introduces a novel chromosome formulation where the
136 genes are not properties of a specific variable but weights for the probability to move in a given direction
137 in the solution space. This allows the method to occasionally explore different nodal characteristics (at
138 mutation rate μ) while conducting a local neighborhood search.

Simulated Data: Random Networks

To test the efficacy of these optimization heuristics in finding the optimal new network connections they were applied to randomly generated networks that vary in complexity and size. Several types of random graph networks were generated to analyze the efficacy of the optimization heuristics for systems with different topologies which are generally representative of naturally occurring and built systems: (1) Erdös-Rényi networks, (2) Watts-Strogatz networks, (3) Barabási and Albert networks, (4) Klemm and Eguílez networks, (5) Delaunay triangulation networks, and (6) Voronoi diagrams. Erdös-Rényi random networks are constructed by randomly creating connection between pairs of nodes with a probability (Erdös and Rényi, 1959). These networks, even though they have random connections, consistently have short average path lengths and irregular connections, both of which are well found in natural systems. The Watts-Strogatz networks also have random connections but the networks also form clusters, another feature commonly found in real-world networks (Watts and Strogatz, 1998). The Barabási-Albert model produces random structures with a small number of highly connected nodes, 'hubs', which are observed in numerous types of networks (Barabási and Albert, 1999; Albert and Barabási, 2002). Klemm and Eguílez networks have random connections, clusters, and hubs (Klemm and Eguílez, 2002). See Supplemental Information Figure S.1 (A) – (D) and Prettejohn et al. (2011) for the algorithms used to generate these networks.

We also introduce two novel types of random planar network versions of the Voronoi diagram and the Delaunay triangulation (Supplemental Information Figure S.1 (E) and (F)). Planarity is particularly important in many fields and networks generated from Voronoi diagrams and Delaunay triangles have been used in spatial health epidemiology (Johnson, 2007), transportation flow problems (Steffen and Seyfried, 2010; Pablo-Martì and Sánchez, 2017), terrain surface modeling (Floriani et al., 1985), telecommunications (Meguerdichian et al., 2001), computer networks design (Liebeherr and Nahas, 2001), and hazard avoidance systems in autonomous vehicles (Anderson et al., 2012). Delaunay triangulation maximizes the minimum angles between three nodes to generate planar graphs with consistent network characteristics while Voronoi diagrams, the dual of a Delaunay triangulation, are composed of points and cells such that each cell is closer to its point than any other point (Delaunay, 1934). To modify these edges are removed from network nodes randomly based on their distance from the focal node with probability

$$p_R \cdot \frac{\max(d(i,F), d(j,F))}{\max_k d(k,F)}, \tag{9}$$

where p_R is the removal probability and weighted by the normalized edge distance from the focal node. When edges are randomly removed from the connected Delaunay network or Voronoi network, with weights given by node distance from a focal node, these networks display some of the properties similarly found in the networks mentioned above, such as complexity and randomness. Yet, these networks have the added component of being planar and having edge weights that can be framed as physical distances.

To compare the efficacy of different optimization methods for different network topologies, identifying the best set of parameters are critical. Parameter values were selected for each type of random network to ensure network complexity (Supplemental Information Table S.6 summarizes the parameters which were used in the analysis). Variation in network size was also explored and the most connected node in each network was selected as the focal node. Uniformly randomly generated edge weights in [0,1] were used for the network distances and the threshold distance was set to ensure that half of the nodes were initially within the distance to the focal node. The costs and benefits were normalized using the ranges from the exhaustive search routine as a benchmark to compare the results from the different optimization methods.

Empirical Data: Transportation Case Study

To complement the random network analysis, the network connectivity optimization methods were applied to a study of urban transportation planning. Network connectivity optimization methods were used to evaluate the potential costs and benefits of increased thoroughfare connectivity for student walking or biking to school. It is assumed that expanding the connectivity around a school would allow for more households, and students, to be included within the walking distance to the school. If more students actively commute to school, this reduces the busing costs for the school system and increases the health and academic achievement of the students (Centers for Disease Control and Prevention, 2010). Yet, streets have associated land, construction, and maintenance costs that are primarily dependent on their length.

Networks composed of street edges and residence nodes around several schools from a representative US school system were used for the analysis. Ten suburban and rural schools from Knox County, TN, were

selected for the analysis, including seven elementary and three middle, that would benefit the most from increased thoroughfare connectivity, i.e., had the most students within the Euclidean walking distance but not the network distance to the school (characteristics of these schools are provided in the Supplemental Information Table S.3). Urban schools were not used since the street connectivity around the schools was significantly high and the from additional thoroughfares would be low. Residential parcels and street networks were provided by the Knoxville-Knox County Metropolitan Planning Commission and the residential parcels were converted to nodes and placed on the nearest street edge. The residences within 1 mile and 1.5 miles, for the elementary schools and the middle schools respectively, are considered close nodes while the nodes outside of these distances were classified as distant nodes (see Figure 2). The school networks do not generally display the characteristics of complex network: they had low average degree, large path lengths, and were not efficient, yet they did have power distributions of connectivity with few intersections having a large number of street connections (see Supplemental Information Table S.3). The networks were evaluated with each optimization method to maximize the number or close residences connected to the school and minimize the distance of the new thoroughfare. The costs and benefits of these street connections were normalized using the ranges from the exhaustive search routine as a benchmark to compare the results from the different optimization methods.
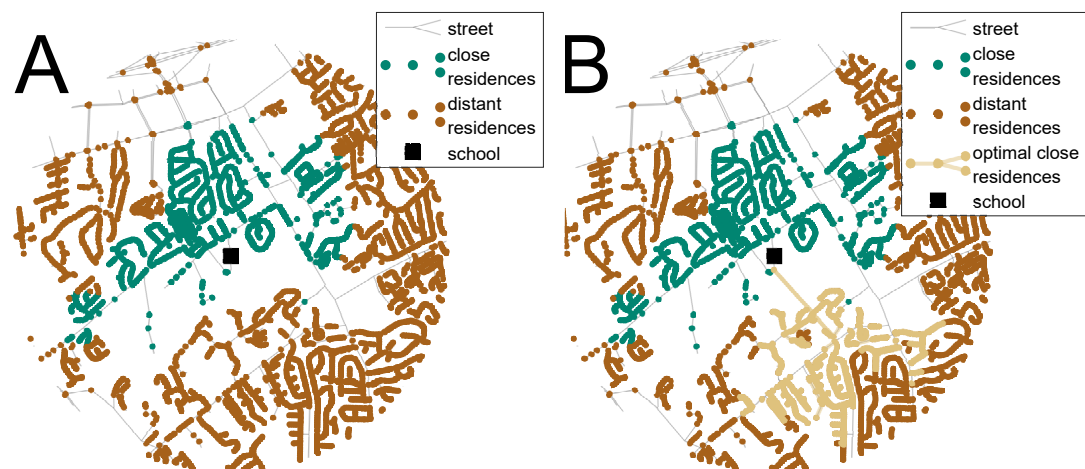
Figure 2. Results of the transportation case study used for the analysis. A network of streets and residences around a school is shown in (A) and with the optimal new walking connection in (B). The red nodes represent the distant residences, i.e., the residences within the 1-mile Euclidean walking distance to the school but not the 1-mile street network walking distance, the green nodes are the close residences within the street network school walking distance, and the black square represents the school. The orange line is the optimal new walking connection that maximizes the number of additional residences (orange nodes) and minimizes the length of the new connection.

Empirical Data: Social Network Analysis

The topology of a network and the management of its system can improve information flow and have been shown to help counter the negative effects of social and environmental crises (Helbing et al., 2015). Specifically, increased social network connectivity can reduce the time information spreads among its members. For example, quickly notifying those nearby and unaware of an active shooter event can save lives. Using a set of social network data coupled with location data, we evaluated the heuristics to optimize social network connectivity for those nearby a specific location.

The data set used for this analysis was from Gowalla, a location-based social networking website where users share their locations by checking-in, and this data set was collected and published by Cho et al. (2011). This social network is undirected and consists of 196,591 nodes (members) and 950,327 edges (social connections). There is a total of 6,442,890 check-ins of these members over the period of February 2009 to October 2010 (network characteristics are provided in the Supplemental Information Table S.4). For the analysis we simulated 100 crisis incidents at a highly populated urban place, Grand

Central Terminal in New York City (NY). Grand Central Terminal was selected as the location of the simulated events as it is a major transportation hub located in the center of the city that serves over a million commuters and visitors daily. New York City is also a major metropolis with tens of thousands of Gowalla users present in the data set and the city has a history of incidents, such as terror attacks. Ten dates were selected at random and for each date ten times were randomly selected between 1200 and 1800 local to simulate a crisis event (see Figure 3).

The social network problem was formulated such that an event occurred at the location (Grand Central Station) and members who were within 0.5-miles of the event should be notified. The members within Grand Central Station are automatically notified of the event whereas those outside can be notified through their online social network if a member in their network is aware. The event is considered to be serious enough that those members aware of it will share the information on the social network. New connections are evaluated based on their number of additional nearby members who become aware of the event.
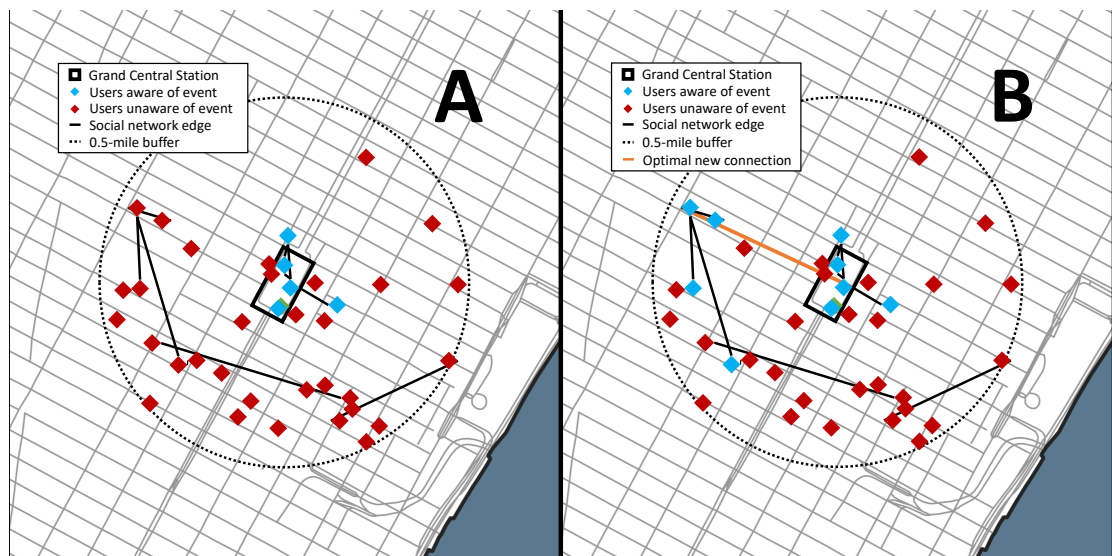
Figure 3. Results of the Gowalla social network case study used for the analysis. For simplicity the network shown is the users with a check-in within 1 hour prior to the event and nearby (within 0.5-mile) the event location, Grand Central Station (NYC). (A) the network prior during the event and (B) with the optimal connection. The red nodes represent the users within 0.5-mile who are unaware of the event, the green nodes are the users aware of the event, the black line represents. The orange line is the optimal new social network connection that maximizes the number of additional nearby users aware of the event.

PERFORMANCE OF THE ALGORITHMS

Several finding are worthy to note regarding the performance of heuristic algorithms used in the analysis. First, there were consistent nonlinear relationships between the nodal characteristics and the quality of the solutions for each type of random network and the school networks (see Figure 4). There was also significant variation for which nodal characteristics were correlated with the quality of the solution across networks (see Table 3). Among those, the distance between the close node and the focal node and the distance between the distant node and the close node were most often highly correlated with the quality of the solution across networks. The centrality measures were inconsistently related to the solution quality for the random networks yet were related to the optimal solutions for the social networks.

The results of the termination times and the optimal solutions deviations from the optimization heuristics applied to the random networks are summarized in Figure 5 and Supplemental Information Figures S.1 and S.2. The hill climbing method was consistently faster for all of the networks, yet had the largest cost and benefit deviations. Simulated annealing and the genetic algorithm had similar termination times, but the genetic algorithm was consistently superior to all of the other methods in approaching the optimal solution. The results from the application of the optimization heuristics applied to the ten school networks are shown in Figure 5 (E) and (F). The times to termination for each heuristic according to
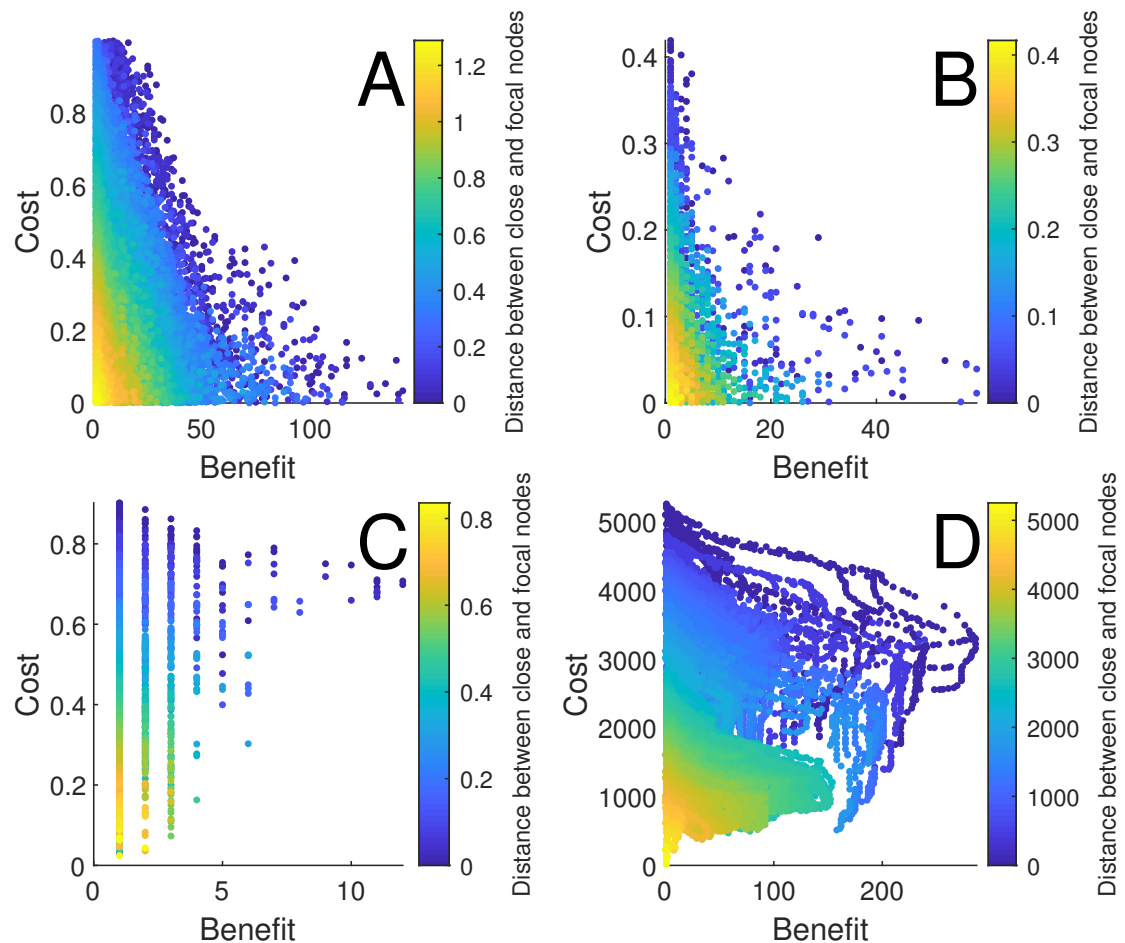
Figure 4. The relationship between the distances of the close node and the focal node with the costs and benefits for each solution for different networks. Figure (A) shows the relationship for a Watts-Strogatz network with $N = 500$, (B) a Barabási and Albert network with $N = 500$, (C) a Delaunay network with $N = 500$, and (D) a suburban school transportation network ($N \approx 4000$). Each point represents a connection between a distant and close node, where the cost is the standardized length of the connection and the benefit is the number of new nodes within the distance to the focal node or school.

237 network size consistently followed the following pattern: ES>SA>HCVN>GA>HCS>HC. The genetic
238 algorithm clearly outperformed the other heuristics, followed by simulated annealing, in terms of cost and
239 benefit deviations (see Figure 5 (B), (D), and (F)).

DISCUSSION, FUTURE WORK, AND LIMITATIONS

241 The local network connectivity problem introduced in this study is relevant to a wide range of applications
242 and is nontrivial as the number of potential solutions can become large even for small systems. This class
243 of combinatorial optimization problem highlights the difficulty in determining local search routines a
244 priori. When the exhaustive search routine was applied to random networks and the real-world networks,
245 the optimal solutions were found to be related to nodal characteristics, which entails a great complexity
246 to find optimal solutions. Therefore, the heuristics employed to reduce the computational costs utilized
247 nodal characteristics to search for solutions. Yet, these nodal characteristics were nonlinearly related
248 to the solutions. Given the example networks, it should be noted that distance to the focal node was
249 consistently related to the quality of the solution as this lowers connectivity length costs, while centrality
250 was intermittently correlated with solution quality it provides greater benefit through more connections.
251 Aside from the distance characteristic, these nodal characteristics also varied in their correlation (sign

²⁵² and magnitude) with solution quality for different types of networks. This makes it difficult to exclude
²⁵³ or prioritize specific nodal characteristics for local network connectivity optimization heuristics. This
²⁵⁴ could arise from the four following issues: (a) the curse of dimensionality, i.e., large sparse subspaces in
²⁵⁵ the solution space; (b) the nodal characteristics are highly correlated with each other; (c) outliers; and
²⁵⁶ (d) the nodal characteristics are heterogeneous across the network. Results from the street networks in
²⁵⁷ the transportation case-study found that the clustering coefficient was a poor measure due to the lack of
²⁵⁸ triplets in the networks.

²⁵⁹ The optimization heuristics save computational time but vary considerably in their ability to find (near)
²⁶⁰ optimal solutions. The stochastic hill climbing search was not effective due to the large neighborhood
²⁶¹ search space explored. In our experiment, the number of solutions checked at each iteration is > 300
²⁶² and resulted in a skewed probability distribution of objective values favoring the selection of low values.
²⁶³ This degraded the efficiency of the method resulting in the selection of poor solutions. The variable
²⁶⁴ neighborhood search method was similarly not reliable because of the significantly large neighborhood
²⁶⁵ search space (the number of possible solutions explored at a given iteration could be $> 5,000$), and had
²⁶⁶ intermediate results with cost and benefit deviations. The simulated annealing heuristic consistently took
²⁶⁷ longer to converge than the other optimization methods from the exploration of suboptimal solutions prior
²⁶⁸ to moving towards better solutions, yet it was able to converge to values close to the optimal solution.

²⁶⁹ The computational costs and the variance in the importance of nodal characteristics for the random
²⁷⁰ networks and real-world systems highlights the need for a heuristic that is able to quickly and effectively
²⁷¹ explore the solution space without getting stuck in a local optimum. The genetic algorithm provided in this
²⁷² work offers a solution to these issues and outperformed the other algorithms in terms of the consistently
²⁷³ higher solution precision and accuracy. The genetic algorithm is able to dynamically reduce the size
²⁷⁴ of the neighborhood search space and what variables to analyze. This reduction in the local solution
²⁷⁵ search space allows the genetic algorithm presented here to converge on solutions near the optimal in a
²⁷⁶ timely fashion. The heuristic is also able to compare solutions from distant search spaces with nonzero
²⁷⁷ probability, thereby avoiding local optimums. The experiments indicate the power of biologically inspired
²⁷⁸ algorithms to effectively explore multidimensional spaces (commonly found in natural systems) and their
²⁷⁹ potential use in a wide variety of disciplines, including the specific applications for planning and crisis
²⁸⁰ management presented above. The combinatorial optimization techniques employed here to identify
²⁸¹ and evaluate new street connections can also complement the optimization approaches used for other
²⁸² transportation planning problems, such as greenway planning (Linehan et al., 1995), bus stop locations
²⁸³ (Ibeas et al., 2010; Delmelle et al., 2012), and health care accessibility (Gu et al., 2010).

²⁸⁴ There are several research directions from these proposed methods. Application of these methods and
²⁸⁵ heuristics can be tested on multi-level networks, such as telecommunication systems, higher dimensional
²⁸⁶ real-world networks (transportation networks with elevation), directed networks, and additional planar
²⁸⁷ random networks (e.g., Gabriel graphs) and this should be conducted. Different distance measures to the
²⁸⁸ focal node, such as the Hamming distance, could also be evaluated for different applications, and other
²⁸⁹ real world examples should be used for analysis. The methods presented here simplified the costs and
²⁹⁰ did not account for many real-world barriers that may restrict optimal new connections found through
²⁹¹ the heuristics. For example, the transportation case study did not include legal considerations, such as
²⁹² right-of-way, or physical barriers, such as highways or rivers. Furthermore, the methods presented here do
²⁹³ not evaluate whether the new connections intersect existing edges, as in existing transportation networks,
²⁹⁴ and attempts to incorporate such a feature resulted in unrealistic computational times.

ACKNOWLEDGMENTS.

²⁹⁶ We would like to thank Alex Zendel (GIS Analyst at the Knoxville-Knox County Metropolitan Planning
²⁹⁷ Commission) for providing the street networks and residential data around the schools.

REFERENCES

²⁹⁹ Albert, R. and Barabási, A. (2002). Statistical mechanics of complex networks. *Reviews of Modern*
³⁰⁰ *Physics*, 74(1):47–97.
³⁰¹ Alenazi, M. J. F., Çetinkaya, E. K., and Sterbenz, J. P. G. (2014). Cost-efficient algebraic connectivity
³⁰² optimisation of backbone networks. *Optical Switching and Networking*, 14(2):107–116.

303   Anderson, E. J. and Ferris, M. C. (1994). Genetic algorithms for combinatorial optimization: The
304   assembly line balancing problem. *ORSA Journal on Computing*, 6:161–173.
305   Anderson, S. J., Karumanchi, S. B., and Iagnemma, K. (2012). Constraint-based planning and control. In
306   *2012 IEEE Intelligent Vehicles Symposium*, pages 383–388. IEEE.
307   Auerbach, J., Fitzhugh, E. C., and Zaviska, E. (2021). Impacts of small changes in thoroughfare
308   connectivity on the potential for student walking. *Journal of Urban Planning and Development*.
309   Barabási, A. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–
310   512.
311   Barrat, A., Barthélemy, M., Pastor-Satorras, R., and Vespignani, A. (2004). The architecture of complex
312   weighted networks. *Proceedings of the National Academy of Sciences USA*, 101(11):3747–3752.
313   Bavelas, A. (1950). Communication patterns in task-oriented groups. *The Journal of the Acoustical*
314   *Society of America*, 22:725.
315   Branas, C. C., MacKenzie, E. J., Williams, J. C., Schwab, C. W., Teter, H. M., Flanigan, M. C., Blatt,
316   A. J., and ReVelle, C. S. (2005). Access to trauma centers in the united states. *Journal of the American*
317   *Medical Association*, 293(21):2626–2633.
318   Brimberg, J. and Hodgson, J. M. (2011). Heuristics for location models. In Eiselt, H. A. and Marianov,
319   V., editors, *Foundations of Location Analysis*, chapter 15, pages 335–355. Springer.
320   Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer*
321   *Networks and ISDN Systems*, 30(1-7):107–117.
322   Centers for Disease Control and Prevention (2010). The association between school based physical
323   activity, including physical education, and academic performance. Technical report, U.S. Department
324   of Health and Human Services.
325   Cho, E., Myers, S. A., and Leskovec, J. (2011). Friendship and mobility: User movement in location-
326   based social networks. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge*
327   *discovery and data mining*, pages 1082–1090.
328   Delaunay, B. (1934). Sur la sphère vide. *Bulletin de l'Académie des Sciences de l'URSS, Classe des*
329   *sciences mathématiques et naturelles*, 6:793–800.
330   Delmelle, E., Shuping, L., and Murray, A. (2012). Identifying bus stop redundancy: A gis-based spatial
331   optimization approach. *Computers, Environment and Urban Systems*, 36:445–455.
332   Demaine, E. D. and Zadimoghaddam, M. (2010). Minimizing the diameter of a network using shortcut
333   edges. In *Proceedings of the 12th Scandinavian Symposium and Workshop on Algorithm Theory*
334   *(SWAT) (Lecture Notes in Computer Science)*, volume 6139, pages 420–431. Springer.
335   Donoso, Y. and Fabregat, R. (2007). *Multi-Objective Optimization in Computer Networks Using Meta-*
336   *heuristics*. Auerbach Publications.
337   Erdös, P. and Rényi, A. (1959). On random graphs. *Publicationes Mathematica*, 6:290–297.
338   Eubank, S., Guclu, H., Kumar, V. S. A., Marathe, M. V., Srinivasan, A., Toroczkai, Z., and Wang, N.
339   (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429:180–184.
340   Floriani, L. D., Falcidieno, B., and Pienovi, C. (1985). Delaunay-based representation of surfaces defined
341   over arbitrarily shaped domains. *Computer Vision, Graphics, and Image Processing*, 32(1):127–140.
342   Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41.
343   Gavrilets, S., Auerbach, J., and van Vugt, M. (2016). Convergence to consensus in heterogeneous groups
344   and the emergence of informal leadership. *Scientific Reports*, 6.
345   Golden, B. L. and Skiscim, C. C. (1986). Using simulated annealing to solve routing and location
346   problems. *Naval Research Logistics Quarterly*, 33(2):261–279.
347   Greiner, R. (1992). Probabilistic hill-climbing: Theory and applications. In *Proceedings of the CSCSI-92*,
348   pages 60–67. Morgan Kaufmann Publishers, Inc.
349   Gu, W., Wang, X., and McGregor, S. E. (2010). Optimization of preventive health care facility locations.
350   *International Journal of Health Geographics*, 9(17).
351   Helbing, D., Brockmann, D., Chadefaux, T., Donnay, K., Blanke, U., Woolley-Meza, O., Moussaid, M.,
352   Johansson, A., Krause, J., Schutte, S., and Perc, M. (2015). Saving human lives: What complexity
353   science and information systems can contribute. *Journal of Statistical Physics*, 158:735–781.
354   Ibeas, A., dell'Olio, L., Alonso, B., and Sainz, O. (2010). Optimizing bus stop spacing in urban areas.
355   *Transportation Research Part E: Logistics and Transportation Review*, 46:446–458.
356   Jaramillo, J. H., Bhadury, J., and Batta, R. (2002). On the use of genetic algorithms to solve location
357   problems. *Computers and Operations Research*, 29:761–779.

Jiang, Z., Liang, M., and Guo, D. (2011). Enhancing network performance by edge addition. *International Journal of Modern Physics C*, 22(11):1211–1226.

Johnson, D. S., Aragon, C. R., McGeoch, L. A., and Schevon, C. (1989). Optimization by simulated annealing: An experimental evaluation; Part 1, Graph partitioning. *Operations Research*, 37(6):865–892.

Johnson, S. (2007). *The Ghost Map: The Story of London's Most Terrifying Epidemic-and How It Changed Science, Cities, and the Modern World*. Riverhead Books.

Khafa, N. A. and Jalili, M. (2019). Optimization of synchronizability in complex spatial networks. *Physica A: Statistical Mechanics and its Applications*, 514(15):46–55.

Kim, K., Dean, D., Kim, H., and Chun, Y. (2016). Spatial optimization for regionalization problems with spatial interaction: a heuristic approach. *International Journal of Geographic Information Science*, 30(3):451–473.

Kirkpatrick, S. (1984). Optimization by simulated annealing: Quantitative studies. *Journal of Statistical Physics*, 34(5/6):975–986.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220:671–680.

Klemm, K. and Eguílez, V. M. (2002). Growing scale-free networks with small-world behavior. *Physical Review E*, 65.

Liebeherr, J. and Nahas, M. (2001). Application-layer multicast with Delaunay triangulations. In *GLOBECOM'01. IEEE Global Telecommunications Conference*, pages 1651–1655. IEEE.

Linehan, J., Gross, M., and Finn, J. (1995). Greenway planning: developing a landscape ecological network approach. *Landscape and Urban Planning*, 33(1-3):179–193.

Meguerdichian, S., Koushanfar, F., Qu, G., and Potkonjak, M. (2001). Exposure in wireless Ad-Hoc sensor networks. In *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking*, MobiCom '01, pages 139–150. Association for Computing Machinery.

Meyerson, A. and Tagiku, B. (2009). Minimizing average shortest path distances via shortcut edge addition. In Dinur, I., Jansen, K., Naor, J., and Rolim, J., editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 272–285. Springer.

Mladenović, N., Brimberg, J., Hansen, P., and Moreno-Pérez, J. A. (2007). The p-median problem: A survey of metaheuristic approaches. *European Journal of Operational Research*, 179:927–939.

Mladenović, N. and Hansen, E. (1997). Variable neighborhood search. *Computers and Operations Research*, 24(11):1097–1100.

Newman, M. E. J. (2008). Mathematics of networks. In Blume, L. and Burlauf, S., editors, *The New Palgrave Encyclopedia of Economics*. Palgrave Macmillan, Basingstoke, 2nd edition.

Oliver, I. M., Smith, D. J. D., and Holland, R. C. J. (1987). Study of permutation crossover operators on the traveling salesman problem. In *Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms and their application*, pages 224–230. MIT.

Pablo-Martì, F. and Sánchez, A. (2017). Improving transportation networks: Effects of population structure and decision making policies. *Scientific Reports*, 7.

Prettejohn, B., Berryman, M., and McDonnell, M. (2011). Methods for generating complex networks with selected properties for simulations: a review and tutorial for neuroscientists. *Frontiers in Computational Neuroscience*, 5.

Resende, M. G. C. and Pardalos, P. M., editors (2006). *Handbook of Optimization in Telecommunications*. Springer.

Russell, S. J. and Norvig, P. (2004). *Artificial Intelligence: A Modern Approach*. Prentice Hall.

Schrijver, A. (2002). On the history of the transportation and maximum flow problems. *Mathematical Programming*, 91(3):437–445.

Steffen, B. and Seyfried, A. (2010). Methods for measuring pedestrian density, flow, speed and direction with minimal scatter. *Physica A: Statistical Mechanics and its Applications*, 389(9):1902–1910.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, 393:440–442.

Wu, F., Huberman, B. A., Adamic, L. A., and Tyler, J. R. (2004). Information flow in social groups. *Physica A: Statistical Mechanics and its Applications*, 337(1-2):327–335.
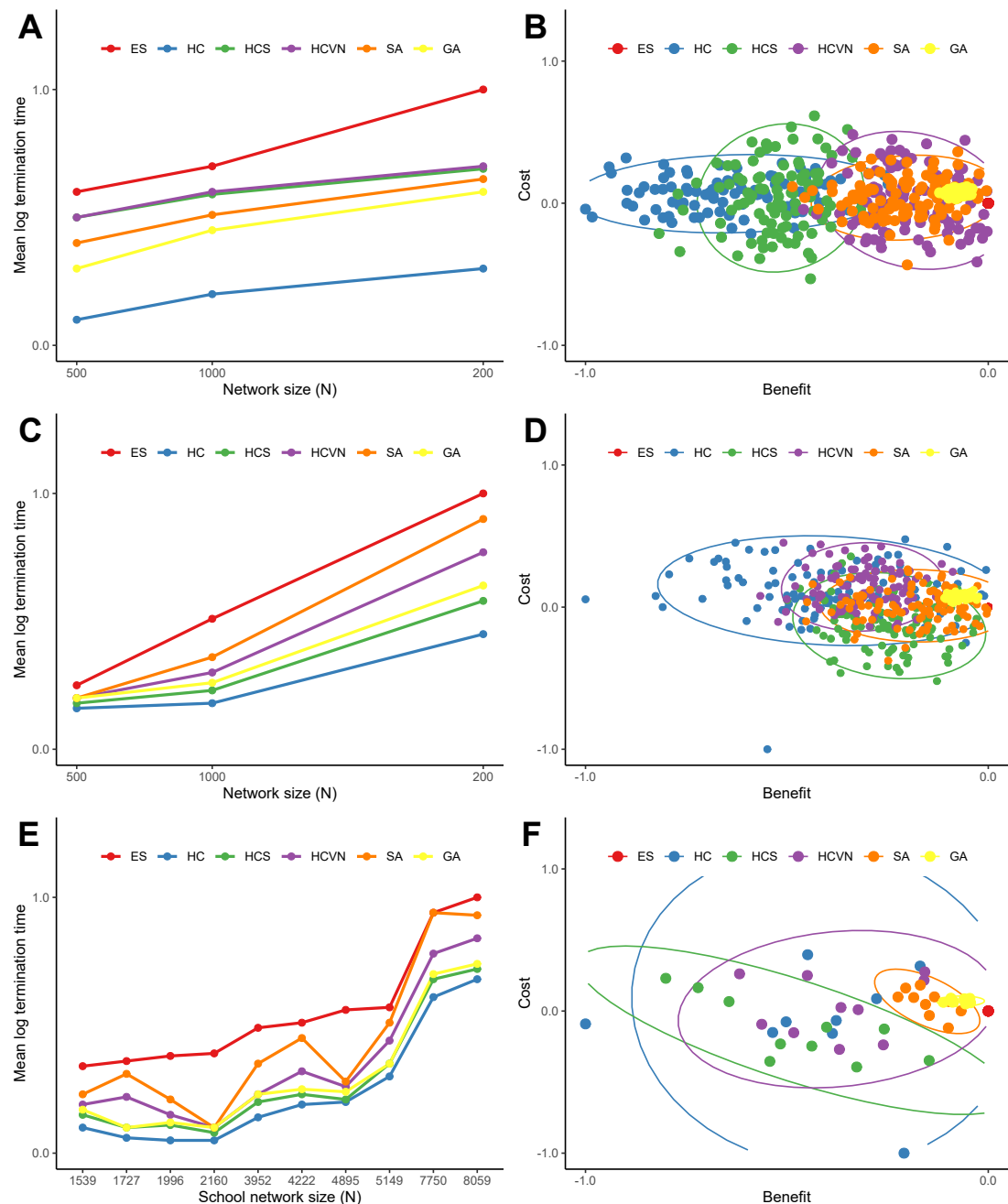
Figure 5. The termination times (Figures A, C, and E) and the differences between the heuristic solution and the global optimal solution costs and benefits (Figures B, D, and F) for the heuristics applied to a sample of random networks and school networks: (A) and (B) Erdös-Rényi networks, (C) and (D) Delaunay networks, and (E) and (F) the ten school networks. For the random networks (Figures A and C) the termination times were averaged over 1000 networks for network sizes of 500, 1000, and 2000 nodes. The times were normalized by the exhaustive search time and log transformed. For the random network cost and benefit differences (Figures B and D) were drawn from a sample of 100 networks with 1000 nodes. The costs were the total lengths of the new connections and the benefits were the number of additional nodes for the solution. The costs and benefits were normalized with the global optimal solution from the exhaustive search, and a longer connection length is a positive cost difference whereas a shorter connection is a negative cost difference. For visualization, the 95-percent confidence ellipses drawn from the Hotelling's T^2 statistic were included. The heuristic acronyms are the following: exhaustive search (ES), hill climbing (HS), stochastic hill climbing (HCS), hill climbing with variable neighborhood (HCVN), simulated annealing (SA), and genetic algorithm (GA).

Symbol	Definition
v	Network node
e	Network edge
N	Number of nodes in a given network, $N = \sum_i v_i$
A	Network adjacency matrix
a_{ij}	Adjacency matrix element ij
F	Focal node
$d(i, j)$	Network distance between nodes i and j
D	Threshold distance from focal node
N_C	Set of close nodes, $N_C \subset N$
N_D	Set of distant nodes, $N_D \subset N$
$N_{i,j}^C$	Set of nodes that are now close after a new connection between nodes i and j
L_F	Average path length to the focal node
$C(i, j)$	Cost of the new connection
$B(i, j)$	Benefit of the new connection
α	Cost weight
β	Benefit weight
t	Optimization iteration
O^t	Optimal solution for iteration t
O^*	Optimal solution
M	Set of long-term memory solutions
C_i^D	Degree centrality of node i
C_i^C	Closeness centrality of node i
σ_{ij}	Shortest path between nodes i and j
$\sigma_{jk}(i)$	Shortest path between nodes j and k that includes node i
C_i^B	Betweenness centrality of node i
λ	Eigenvalue
x_i	Eigenvector
C_i^E	Eigenvector centrality of node i
α_P	Attenuation factor
C_i^P	Pagerank centrality of node i
η	Variable neighborhood size
μ	Genetic algorithm mutation rate
s	Genetic algorithm selection coefficient
P	Population of solutions for the genetic algorithm
$f(i, j)$	Genetic algorithm fitness function
ε_B	Benefit error from heuristic
ε_C	Cost deviation from heuristic
p	Connection probability (Erdös-Rényi graphs and Klemm and Eguílez graphs)
p_W	Rewiring probability (Watts-Strogatz graphs)
k_L	Initial node degree (Watts-Strogatz graphs)
m_0	Initial network size (Barabási and Albert graphs and Klemm and Eguílez graphs)
m	Degree of new nodes (Barabási and Albert graphs)
p_S	node selection probability (Klemm and Eguílez graphs)
p_R	Edge removal probability (Delaunay and Voronoi random graphs)
$\overline{C_D}$	Mean degree of a network
$\overline{L}$	Average path length of a network
c_i^w	Weighted clustering coefficient for node i
w_{ij}	Weight of connection between nodes i and j
$\overline{C}$	Weighted clustering coefficient of a network
$\overline{C_r}$	Weighted clustering coefficient of a completely random network
$\overline{L_r}$	Average path length of a completely random network
γ	Power law exponent
$P(n)$	Degree distribution
E	Efficiency of a network
E_r	Efficiency of a completely random network
E_G	Global efficiency of a network

Table 1. List of symbols and their definitions.

Local search selection criteria	
distance from focal node $d(i,F)$	degree centrality $C_i^D = \sum_j A(i,j)$
closeness centrality $C_i^C = 1/\sum_j d(i,j)$	betweenness centrality $C_i^B = \sum_{j \neq i \neq k} \dfrac{\sigma_{jk}(i)}{\sigma_{jk}}$
eigenvector centrality $C_i^E = \dfrac{1}{\lambda}\sum_j A(i,j)x_j$	pagerank centrality $C_i^P = \alpha_P \sum_j A(i,j)\dfrac{x_j}{\sum_i A(i,j)} + \dfrac{1-\alpha_P}{N}$

$$\text{weighted clustering coefficient}$$
$$c_i^w = \frac{1}{(C_i^D - 1)\sum_j a_{ij}w_{ij}}\sum_{j,h}\frac{(w_{ij}+w_{ih})}{2}a_{ij}a_{ih}a_{jh}$$

Table 2. Node characteristics used for the neighborhood search and their formulations.

			Network							
			ER	WS	BA	KS	DT	VD	Schools	Social Network
Nodal characteristics	close node	$d(j,F)$	-0.08	**-0.52**	**-0.37**	**-0.48**	-0.28	**-0.41**	**-0.30**	**0.11**
		C_j^D	0.14	0.08	0.05	0.09	-0.01	0.02	0.02	0.09
		C_j^C	0.10	**0.26**	0.05	**0.15**	0.15	0.15	**0.18**	0.09
		C_j^B	0.14	0.15	0.05	0.04	-0.02	0.03	0.06	0.05
		C_j^E	0.12	0.11	0.05	0.09	0.01	0.08	-0.01	-0.07
		C_j^P	0.14	0.07	0.05	**0.10**	-0.03	-0.01	-0.00	-0.08
		c_j^w	0.00	-0.03	0.02	0.00	0.14	0.03	*	-0.04
	distant node	$d(i,F)$	**-0.46**	**-0.18**	**-0.12**	-0.09	-0.07	-0.08	-0.04	-0.07
		C_i^D	0.17	0.03	**0.08**	0.05	**0.29**	**0.30**	**0.08**	**-0.30**
		C_i^C	0.15	-0.01	0.07	0.00	**0.35**	0.21	0.06	**0.26**
		C_i^B	**0.19**	0.03	0.07	0.03	0.20	0.20	0.07	-0.09
		C_i^E	**0.21**	0.02	0.07	0.06	-0.16	-0.01	-0.03	-0.10
		C_i^P	0.17	0.01	0.06	0.07	**0.30**	**0.22**	0.01	0.05
		c_i^w	0.08	-0.01	0.03	0.03	0.12	0.12	*	-0.08

Table 3. Mean correlation coefficients for the nodal characteristics and the solution benefits for the experimental networks. The three coefficients with the largest magnitude are highlighted in bold for each network type. (*) There was no variation in clustering coefficients as triplets were not common in the street networks.