

Identifying effective sink node combinations in spacecraft data transfer networks

Ruaridh Clark (✉ ruaridh.clark@strath.ac.uk)

University of Strathclyde

Ciara Nora McGrath

University of Strathclyde

Malcolm Macdonald

University of Strathclyde

Research Article

Keywords: flow network, space system, consensus dynamics

Posted Date: March 2nd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1405393/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

Identifying effective sink node combinations in spacecraft data transfer networks

Ruaridh A Clark^{1*}, Ciara N McGrath¹ and Malcolm Macdonald¹

*Correspondence:

ruaridh.clark@strath.ac.uk

¹Department of Electronic and Electrical Engineering, University of Strathclyde, George Street, Glasgow, UK

Full list of author information is available at the end of the article

Abstract

Complex networks are emerging in low-Earth-orbit as the communication architectures of inter-linked space systems. These data transfer networks vary based on spacecraft interaction with targets and ground stations, which respectively represent source and sink nodes for data flowing through the network. We demonstrate how networks can be used to identify effective sink node selections that, in combination, provide source coverage, high data throughput, and low latency connections for intermittently connected, store-and-forward space systems. The challenge in this work is to account for the changing data transfer network that varies significantly depending on the ground stations selected – given a system where data is downlinked by spacecraft at the first opportunity. Therefore, *passed-on* networks are created to capture the redistribution of data following a sink node's removal from the system, a problem of relevance to traffic management in a variety of flow network applications. Modelling the system using consensus dynamics, enables sink node selections to be evaluated in terms of their source coverage and data throughput. While restrictions in the depth of propagation when defining passed-on networks, ensures the optimisation implicitly rewards lower latency connections. This is a beneficial by-product for both space system design and store-and-forward data networks in general. The passed-on networks also provide an insight into the relationship between sink nodes, with eigenvector embedding-based communities identifying sink node divisions that correspond with differences in source node coverage.

Keywords: flow network; space system; consensus dynamics

1
2

3 Introduction

4 Historically, satellite constellations were composed of a few large spacecraft that pro-
5 duced simple, grid-like, communication network topologies Dietrich (1997); Keller

6 and Salzwedel (1996); Pratt *et al.* (1999). In contrast, new small-satellite constel-
7 lations present as complex data transfer networks due to the variety of orbital
8 positions, as a result of a reliance on ad-hoc launch opportunities. This presents
9 a challenge for operators to efficiently select, or locate, ground stations that can
10 suitably service their constellation. This paper demonstrates how holistic assess-
11 ment of these complex networks can provide an analytical approach to geographical
12 ground station selection - a highly combinatorial problem. Such an approach opens
13 up the potential for agile and responsive space systems that can be adapted by
14 altering their connectivity to the ground, rather than relying on costly and limited
15 spacecraft manoeuvring capabilities. While the developed approach is shown to be
16 effective for the specific challenge of space system analysis, it is expected that the
17 presented methods could be effectively applied to a range of similar problems seen
18 in, for example, traffic flow systems (see Nath and Dhamala (2018)) and wireless
19 sensor networks (see Kim *et al.* (2005); Safa *et al.* (2014)).

20 Data transfer is a spreading process that Clark *et al.* (2019) showed can be rep-
21 resented by a network in order to detect the relative influence of nodes. A network
22 of averaged contacts over time enables the network's adjacency matrix to provide
23 insights into the major pathways for spread, where Clark and Macdonald (2021)
24 demonstrated this by identifying influential disease spreaders in contact networks.
25 For space system flow networks, where targets are sources of data and ground sta-
26 tions are sinks, Clark *et al.* (2022) detailed how the eigenvectors of the adjacency
27 matrix can reveal the relative influence of ground stations in terms of receiving tar-
28 get data. However, the aggregation of contact times, to approximate data transfer
29 as in Clark *et al.* (2022), limits the applicability of the approach to a system deal-
30 ing with the transfer of discrete data packets – as is the case in many applications
31 including Earth observation and Internet of Things (IoT) services. Since the order
32 in which a spacecraft comes into contact with targets and ground stations plays
33 an important role in determining system performance. To address this challenge,
34 we go beyond the work presented in Clark *et al.* (2022) by proposing an aggre-
35 gated network that accounts for the temporal ordering of contacts. This includes
36 the redistribution of data when a ground station is removed from the system, a
37 necessary step in evaluating an effective subset selection from a set of candidate
38 ground station locations. The redistribution of data provides an estimate for where

39 data will go if a ground station is no longer included in the system, which impacts
40 the relative influence of each sink node within the network.

41 The ground station selection problem is highly combinatorial and with an objec-
42 tive that varies depending on the application. A common objective for data transfer-
43 ring space systems is the reduction of latency; the delay from a spacecraft acquiring
44 data to the receipt of that data on the ground (see Mazzarella et al. (2020)). Along-
45 side latency, target coverage is an important consideration. Targets are defined here
46 as locations on Earth from which data is collected, whether through communication
47 or some other form of sensor acquisition. A network-based approach for ground sta-
48 tion selection is proposed herein, which avoids the need to evaluate the full range of
49 feasible ground station (sink node) selections; this is often an intractable problem,
50 particularly when each assessment requires a detailed and time-consuming data
51 transfer simulation. The network representation proposed herein can be used to ex-
52 plicitly optimise the sink node selection in terms of source node (target) coverage,
53 while the optimisation implicitly rewards lower latency solutions in its estimation
54 of the data transfer network.

55 The majority of systematic ground station selection papers, to date, have fo-
56 cused on large, latency-prioritising constellations that maintain continuous contact
57 between targets and ground stations (referred to as *bent-pipe* systems). Examples
58 of these systems include OneWeb and Starlink, where target-ground station geo-
59 graphical proximity Chen et al. (2021); del Portillo et al. (2019) has been shown
60 to drive ground station placement and *minimum cost, maximum flow* optimisation
61 has been used to define effective inter-satellite link topologies del Portillo et al.
62 (2019). For many other applications involving data collection, latency is an impor-
63 tant but not singular goal of the constellation. Such as *store-and-forward* systems
64 – where spacecraft gather information from one location (e.g. ship AIS beacons or
65 Earth monitoring images) and deliver it to another surface location (referred to as a
66 ground station) – that are the focus of this paper as ground station placement must
67 account for latency, target coverage and data throughput. Additionally, in contrast
68 to fully interconnected bent-pipe systems, the order of connections in the tempo-
69 rally varying topology of the store-and-forward contact network must be considered
70 when determining effective sink node selections.

71 In the past, ground station network design has relied heavily on engineering judge-
72 ment and best practices. Lacoste et al. demonstrated the difficulties in applying
73 best-practices for selecting multiple ground stations Lacoste et al. (2011). They
74 found it difficult to predict the contribution of an additional ground station to an
75 existing set, highlighting the need for combinatorial optimisation methods for the
76 ground station selection problem. An optimised selection of ground station has been
77 proposed by Capelle et al. (2019) for a spacecraft with free-space optical communi-
78 cation, which has communication restricted by cloud cover. The optimisation objec-
79 tive in Capelle et al. (2019) aims to maximise the percentage of data acquired from
80 a single spacecraft. This differs from the target-centric multi-spacecraft problem
81 presented herein, but it does highlight the combinatorial optimisation challenges of
82 the problem and presents both an exhaustive enumeration, similar to that described
83 herein, and a branch-and-bound approach to identify effective subset selections. Tai-
84 loring a ground station selection to a system’s priorities is attractive both as a cost
85 saving measure and as a means to achieve a robust and adaptable system without
86 having to alter the assets in space. With services offering leaseable ground station
87 sites around the globe, this paper presents an approach for space system designers
88 to maximise constellation performance as mission objectives and target priorities
89 change.

90 **Methods**

91 This section describes the pipeline for identifying an effective subset selection of
92 sink nodes. The steps involved are as follows:

- 93 • Propagate the movements of all spacecraft in the Space System Scenarios to
94 create a contact schedule (C).
- 95 • Generating data transfer networks, including a data transfer network (Λ) and
96 a passed-on network ($B[g]$) for every ground station $g \in G$. These networks
97 combine to produce an estimated data transfer network (A) for a given subset
98 selection of ground stations.
- 99 • Identify an Initial eigenvector-based selection of ground stations (sink nodes)
100 using an eigenvector embedding of a ground station relationship network (Γ).

- 101 • Perform an Exhaustive search optimisation based on Consensus dynamics for
102 target coverage, where the objective is to rapidly drive source (target) nodes
103 to consensus under the influence of sink nodes (see Problem definition).

104 Space System Scenarios

105 The space system studied, and the simulation used to create a contact schedule
106 (C), are described in more detail by Clark et al. (2022), but the relevant aspects
107 are summarised here. The space system considered is based on the orbital positions
108 and targets of the Spire Global, Inc. constellation that collects AIS data from ships
109 globally. All 111 spacecraft that as of July 2021 were operated by Spire Global,
110 Inc. are included in this case study, with their Keplerian orbit elements detailed
111 in data set McGrath and Clark (2021). The spacecraft are in differing orbit planes
112 with 74 in sun-synchronous orbits, 22 at approximately 51.6 degrees inclination, 8 at
113 approximately 37 degrees inclination, 4 in near-polar orbits, and 3 in near-equatorial
114 orbits.

115 A representative set of target locations are defined for the case studies, based
116 on data provided by Spire Global, Inc. for the 24-hour period of 11-August-2019
117 14:09 UTC to 12-August-2019 14:08 UTC. This dataset provides the last reported
118 position of all ships detected in this 24-hour window. From this, 250 targets are
119 positioned to approximate the locations of ships worldwide that were tracked from
120 space (rather than via ground-based coastal AIS receivers) with these locations
121 visualised in the Results section (Fig. 3). These 250 targets define the *global targets*
122 scenario, while a sub-set of 16 targets located near the Caribbean are taken as the
123 basis of the *Caribbean* scenario. Twenty ground station sites are considered in this
124 study, with the locations of these sites also visualised in the Results section (Fig. 3).

125 A fixed-step integrator is used to propagate the motion of spacecraft for a defined
126 period of time (T) and time step (τ) to identify contacts (i.e. visible ground stations
127 or targets on the ground). These contacts are collated in a contact schedule (C),
128 which is used to determine the data transfer networks.

129 Generating data transfer networks

130 A data transfer network (Λ) is created to capture the data transactions in the
131 space system, with a set of ground targets, spacecraft in given orbits, and a set of
132 candidate ground stations. The network is populated by propagating the satellites'

133 motion and simulating data transfer in the system for a defined period of time (T),
 134 during which the movement of data packets is monitored. The process of generating
 135 Λ is described in detail in Alg. 1 (in black text), and is summarised as follows:

- 136 • Each spacecraft in the system is assigned a data buffer ($db1$), where data is
 137 inserted when the spacecraft is in contact with a ground target according to
 138 the contact schedule (C).
- 139 • Each packet of data is associated with the target of origin (d) when inserted
 140 into the buffer $db1$.
- 141 • When the spacecraft is in contact with a ground station, packets in the buffer
 142 $db1$ are removed until a downlink/packet removal limit (δ) for a single time
 143 step is reached.
- 144 • For each data packet removal from $db1$, the data transfer network (Λ) is
 145 updated with $\Lambda_{d,n_D+g} = \Lambda_{d,n_D+g} + 1$, where the d is the target of origin, g
 146 is the current ground station (in contact with the spacecraft), and n_D is the
 147 number of targets. Therefore, by the end of the simulation Λ_{d,n_D+g} will equal
 148 the number of packets acquired from d and downlinked to g .

149 In addition to generating the data transfer network (Λ), a passed-on network (B)
 150 must be created for each ground station to estimate where data will be transferred
 151 if that ground station were removed from the system. This allows the importance
 152 of each ground station to be better understood, since not all sink nodes in Λ will be
 153 present in the final subset selected. This process is intertwined with the generation
 154 of Λ and hence is also detailed in Alg. 1 (in blue text), but can be summarised as
 155 follows:

- 156 • Each spacecraft in the system is given a second data buffer ($db2$), which is
 157 populated with dummy data (0 entries) when in contact with targets (i.e.
 158 matching the data inserted into $db1$).
- 159 • In addition to dummy data, a *passed-on* data packet $[d, g]$ is inserted into $db2$
 160 for every data packet d that is removed (i.e. downlinked) from $db1$ for the
 161 same spacecraft, where g is the current ground station (in contact with the
 162 spacecraft).
- 163 • When a spacecraft is in contact with a ground station, the dummy (0 entry)
 164 data is the first to be removed from $db2$ before any of the passed-on data

Algorithm 1: Generating data transfer and passed-on networks

Input: downlink limit (δ), ground station set (G), number of ground stations (n_G), **packet re-insertion limit** (ρ), Simulated time period (T), simulation time step (τ), spacecraft set (S), spacecraft contact schedule (C), target set (D), number of targets (n_D)

Output: data transfer network (Λ), **passed-on networks** (B)

Set $\Lambda \leftarrow \text{zeros}(n_D + n_G, n_D + n_G)$

for $g \in G$ **do**
 Set $B[g] \leftarrow \text{zeros}(n_G, n_G)$

for $s \in S$ **do**
 $db1, db2 \leftarrow$ new data buffer lists ; // data buffers
 $n_{db1} \leftarrow 0, n_{db2} \leftarrow \text{zeros}(n_G)$; // downlink counters
 for $t \leftarrow 0$ **to** T **by** τ **do**
 for $d \in D$ **do**
 if $d \in C[t]$ **then**
 Insert d into $db1$; // data acquired from target
 Insert 0 into $db2$; // dummy data
 for $g \in G$ **do**
 if $g \in C[t]$ **then**
 $\Omega \leftarrow G \in C[t, t + \tau, \dots, t + (\tau \times 9)]$; // set of ground stations soon to be or currently in contact with s
 while ($0 < n_{db1} < \delta$ **and** $db1$ not empty) **or** ($0 < n_{db2} < \delta$ **and** $db2$ not empty) **do**
 for $d \in D$ **do**
 if $0 < n_{db1} < \delta$ **and** $d \in db1$ **then**
 Remove d from $db1$
 $\Lambda_{d, n_D + g} = \Lambda_{d, n_D + g} + 1$; // update Λ
 $n_{db1} = n_{db1} + 1$; // add to downlink count
 Insert $[d, g]$ into $db2$; // passed-on data
 if $0 < \min_{\gamma \in G} (n_{db2}[\gamma]) < \delta$ **then**
 if $0 \in db2$ **then**
 Remove 0 from $db2$
 for $\gamma \in G$ **do**
 $n_{db2}[\gamma] = n_{db2}[\gamma] + 1$; // add to downlink count
 else
 for $\gamma \in G$ **do**
 if $0 < n_{db2}[\gamma] < \delta$ **and** $[d, \gamma] \in db2$ **and** $\gamma \notin \Omega$ **then**
 Remove $[d, \gamma]$ from $db2$
 if packet re-insertion count $< \rho$ **then**
 Insert $[d, g]$ into $db2$; // passed-on data
 Add 1 to packet re-insertion count
 $B[\gamma]_{d, n_D + g} = B[\gamma]_{d, n_D + g} + 1$; // update $B[\gamma]$
 $n_{db2}[\gamma] = n_{db2}[\gamma] + 1$; // add to downlink count

165 packets associated with ground stations. Only once all the dummy data is

166 removed, then the passed-on data packets are removed from $db2$.

- 167 • For each passed-on data packet $[d, \gamma]$ removed from $db2$, while the spacecraft
168 is in contact with ground station g , the entry in the data transfer network
169 $B[\gamma]$ is updated as $B[\gamma]_{d, n_D+g} = B[\gamma]_{d, n_D+g} + 1$, where γ identifies the ground
170 station of origin for the passed-on data. Therefore, by the end of the simulation
171 $B[\gamma]_{d, n_D+g}$ will equal the number of packets that were originally acquired from
172 d , but were passed-on from ground station γ to g .
- 173 • When a passed-on data packet $[d, \gamma]$ is removed from $db2$, a new packet $[d, g]$
174 is inserted into $db2$ that is associated with the current ground station contact
175 (g). The number of times a passed-on data packet is re-inserted back into $db2$
176 is restricted by a packet re-insertion limit (ρ).
- 177 • As with the removal of data from $db1$, the downlink limit is monitored for
178 packets removed from $db2$. However, a separate count of packets removed is
179 maintained for each ground station of origin (γ) for passed-on data.
- 180 • Note that passed-on data packets are not removed from $db2$ if their ground
181 station of origin (γ) is either the current ground station or in close proximity
182 to the ground station γ . This is necessary to avoid the majority of passed-data
183 packets from travelling back and forth between nearby ground stations, and
184 monitored with Ω as detailed in Alg. 1.

185 The *packet re-insertion limit* (ρ) is an important consideration, as this determines
186 the number of times a data packet is passed from one ground station to another.
187 The most accurate passed-on matrices were generated when using a ρ value that is
188 similar to the expected average number of unselected ground stations that will pass
189 on data before it arrives at a selected selection. In this paper we are considering a
190 subset selection of five ground station from a set of 20 candidates, therefore data
191 packets can be estimated to, on average, pass through three ground stations before
192 alighting at a selected station. Given that a significant portion of data packets could
193 pass through more than three ground stations, $\rho = 4$ was applied herein.

194 Estimating data transfer network

195 The difficulty in identifying effective ground station combinations stems from the
196 impact that one selection has on the value of other ground stations in receiving
197 data and covering targets. For example, a ground station (GS1) may be viewed by
198 a spacecraft that has received data from a target (T1). However, it is possible that

199 the data transfer network (Λ) does not report this connection if, for instance, the
 200 spacecraft has already downlinked all of T1's data to other ground stations prior
 201 to overflight of GS1. Therefore, we propose an approach for estimating the data
 202 received by a subset selection of ground stations, using the passed-on networks (B)
 203 to identify where data would go if a ground station was removed. This approach
 204 has been formulated for the analysis of space systems, but such an approach is
 205 generalisable to combinatorial flow network problems, where the removal of a sink
 206 node results in greater traffic arriving at other sinks in the network.

207 To estimate the data received by a subset selection, the data transfer network
 208 (Λ) defined for the full set of ground stations needs to be updated according to the
 209 passed-on networks (B). This process creates an estimated data transfer network
 210 (A) and is detailed in Alg. 2, with the ground station selection represented by a
 211 vector \mathbf{r} where $r_g = 1$ indicates a selected ground station g , and $r_g = 0$ denotes an
 212 unselected ground station. The process involves moving data from each unselected
 213 ground station, in turn, by using the normalised passed-on matrix (K) to determine
 214 where the data goes, before removing data from the ground station's column in the
 215 data transfer network A , and then redistributing the removed data according to K .

Algorithm 2: Estimating data transfer network for a given selection vector

Input: data transfer network (Λ), passed-on networks (B), target set (D), ground station set
 (G), number of data pass iterations (n_{pass}), ground station selection vector (\mathbf{r})

Output: estimated data transfer network (A)

```

 $A \leftarrow \Lambda$ 
for  $g \in G$  do
  for  $d \in D$  do
    for  $g \in G$  do
      if  $\sum_j B[g]_{d,j} > 0$  then
         $s_d \leftarrow \frac{1}{\sum_d B[g]_{d,j}}$ 
      else
         $s_d \leftarrow 1$ 
       $\Phi[g] = \text{diag}(\mathbf{s}) B[g]$  ; // row normalised
for  $n \leftarrow 1$  to  $n_{pass}$  do
  for  $g \in G$  do
     $K = \text{diag}(A_{D,g}) \Phi[g]$  ; // identify data movement
     $A_{D,g} = A_{D,g} - A_{D,g} \times (\max(\mathbf{r}) - r[g])$  ; // remove data
     $A = A + \text{diag}(\max(\mathbf{r}) - r[g]) K$  ; // move data
 $A = \text{diag}(\frac{r}{\max(r)}) A$  ; // only keep data for selected ground stations

```

216 The logic used to determine a suitable *packet re-insertion limit* (ρ) for Alg. 1 is
 217 also relevant for selecting a suitable n_{pass} for Alg. 2. The ρ value determines how
 218 many times a passed-on data packet is re-inserted into the data buffer (*db2*), while
 219 n_{pass} represents the number of times data is moved on from unselected ground
 220 stations when estimating the data transfer network (A). Since data packets can be
 221 estimated to pass through, on average, three ground stations before alighting at
 222 a selected station, then $n_{pass} \geq 3$ could be expected to allow the estimated data
 223 transfer network to capture the majority of redistributed data. As will be discussed
 224 in the Results section, n_{pass} cannot simply be set as a large value to capture all
 225 redistribution of data as this can over-estimate the volume of target data received
 226 by a subset selection of ground stations.

227 Consensus dynamics for target coverage

228 An effective way of evaluating a subset selection of ground stations in terms of tar-
 229 get coverage and data throughput, for a network $G = (V, E)$ of targets and ground
 230 stations, is through the use of consensus dynamics. Specifically consensus leader-
 231 ship, where ground station selections are identified by assessing their ability to lead
 232 targets to consensus – according to the following consensus protocol – when the
 233 connections are defined in the estimated data transfer network (A).

We consider a system where each node v_i has a state $x_i \in \mathbb{R}$ and continuous-time
 integral dynamics, $\dot{x}_i[t] = u_i[t]$ where $u_i \in \mathbb{R}$ is the control input for agent i . The
 linear consensus protocol is

$$u_i(t) = \sum_{j \in N_i} a_{ij}(x_j[t] - x_i[t]) \quad (1)$$

234 and describes how node v_i adjusts its state at time step (t) based on the es-
 235 timated data transfer matrix ($A = [a_{ij}]$) and the node state (x) of its neigh-
 236 bours (N_i). Given this protocol, the state of the network develops according to
 237 $\dot{x}[t] = -Lx[t]$ with the graph Laplacian matrix, L , defined as $L = D - A$ where
 238 $D = \text{diag}(\text{out}(v_1), \dots, \text{out}(v_n))$ is a diagonal matrix composed of the outdegrees of
 239 each node, i.e. $\text{out}(v_i) = \sum_j a_{ij}$.

Given the definitions for the continuous-time integral dynamics and $\dot{x}_i[t]$, the discrete-time agent dynamics are given in Di Cairano et al. (2008) as

$$x_i[t + 1] = x_i[t] + \epsilon u_i[t] \quad (2)$$

240 provided that $0 < \epsilon < \frac{1}{\max_i d_{ii}}$ where d_{ii} is an element of D . The choice of ϵ affects
 241 the number of steps required for nodes to reach convergence, therefore setting $\epsilon =$
 242 $0.999 \times \frac{1}{\max_i d_{ii}}$ allows the number of computational steps to be reduced while still
 243 guaranteeing convergence of the system (see Di Cairano et al. (2008)). Convergence
 244 is defined here as $\bar{x}_i > 0.99 \forall i \in D$, where D is the set of all target (source) nodes,
 245 when $x_j = 1 \forall j \in G$ with G the set of all ground station (sink) nodes.

246 The most effective ground station selections, in terms of target coverage, are those
 247 that achieve the fewest steps until all of the targets reach consensus. Such a selection
 248 would demonstrate a strong connection to all of the targets in the system. If, in
 249 contrast, a selection had no connectivity to a given target then consensus would
 250 never be reached.

251 Problem definition

An objective function is required to optimise the ground station selection. The number of steps to convergence can be used, but it creates a discontinuous search space. Therefore, the mean consensus leadership,

$$m = \frac{\sum_{i \in D} (1 - x_i[t])}{n_D} \quad (3)$$

252 provides a continuous alternative to maximise the mean consensus state of all target
 253 nodes, where n_D is the number of targets (source nodes) and D the set of all targets.
 254 The target (source) nodes states, $x_i[t]$, are evaluated according to Eq. 2 where t is
 255 taken as a point prior to convergence, defined as the closest step to $0.9 \times s_{ref}$
 256 where s_{ref} is the number of steps to convergence. The reference number of steps,
 257 s_{ref} , is defined using the number of steps to convergence required for the Initial
 258 eigenvector-based selection.

259 The optimisation can then be defined as follows,

$$\begin{aligned}
\min \quad & \frac{\sum_{i \in D} (1 - x_i[t])}{n_D} \\
\text{s.t.} \quad & r_g = 1 \quad \forall g \in \Phi \\
& r_g = 0 \quad \forall g \in G \setminus \Phi \\
& \sum_j r_j = n_{select} = |\Phi| \quad \forall j \in G, \quad n_{select} \in \mathbb{Z}^+
\end{aligned} \tag{4}$$

260 where \mathbf{r} is the ground station selection vector, Φ is the subset selection of ground
261 station, G is the set of all ground station candidates and n_{select} the cardinality of
262 the subset Φ .

263 Initial eigenvector-based selection

264 The optimisation of ground station selections is a highly combinatorial problem
265 and as such susceptible to local optima far from the global optimum. This issue
266 is exacerbated by the need to update the data transfer network for every possible
267 selection. We propose an eigenvector embedding-based selection to act as an effective
268 initial selection, providing an alternative to a more exhaustive search. The use of
269 brute-force evaluation of all combinations is often intractable for sufficiently large
270 numbers of candidates and selection sizes.

The relationship of interest, when optimising a system for target coverage, is that between targets and ground stations. However, it is not possible to directly capture this relationship in a static network. Instead a ground station relationship network (Γ) is introduced, based on the passed-on networks B , which details the volume of data that each ground station passes on to every other ground station when removed from the system. While the passed-on networks, $B[g] \quad \forall g \in G$, detail the movement of data from targets to ground station, the network Γ details the connections between ground stations, where

$$\Gamma_{g,\gamma} = \sum_{d \in D} B[g]_{d,\gamma}. \tag{5}$$

271 The Γ network is useful in identifying influential ground stations. This is despite
272 Γ only detailing the relationships between ground stations, since these relationships
273 are a product of connectivity to spacecraft that have collected target data. There-
274 fore, Γ highlights whether ground stations are connected to spacecraft in similar

275 or different orbits. Differing spacecraft orbits result in different target contacts,
276 where these differences lead to different patterns of target coverage. Hence, select-
277 ing ground stations that cover different sets of spacecraft will also likely provide a
278 selection that covers differing communities of targets.

279 The process of ground station selection takes inspiration from work on commu-
280 nities of dynamical influence (CDI), introduced in Clark et al. (2019), that are
281 shown to highlight effective leadership in networks under consensus dynamics. The
282 selection is based on the eigenvectors of Γ , where the dominant eigenvectors (those
283 associated with the largest eigenvalue entries) are used to embed the network in a
284 Euclidean space. The nodes in this space that are furthest from the origin, along the
285 direction of their position vector, are defined as leaders of separate ground station
286 communities. This is assessed by comparing the magnitude of each node's posi-
287 tion vector with the scalar projection onto this vector from all other node position
288 vectors.

289 The explicit objective of the optimisation is to improve target coverage and data
290 throughput, therefore CDI analysis of Γ can facilitate the selection of ground sta-
291 tions. Specifically, an effective combination of ground stations can be expected to
292 involve nodes in multiple different communities to ensure target coverage, while the
293 nodes with the largest first left eigenvector (\mathbf{v}_1) entries are more likely to ensure
294 high data throughput. Therefore, an initial selection composed of ground stations
295 from different CDIs, each with the largest \mathbf{v}_1 entry will provide a good initial guess.

296 Exhaustive search optimisation

297 An optimal selection of ground stations, in terms of convergence to consensus for
298 the space system modelled using consensus dynamics, can be obtained by simulat-
299 ing all subset combinations from a set of candidates. For the scenarios explored in
300 this paper that involves simulating all combinations of five ground stations from
301 20 possible options (15503 combinations in total). This is a computationally in-
302 tensive process that required approximately 10 days (60 seconds per simulation)
303 computation time for the global targets scenario on a desktop machine – Intel Xeon
304 Processor with 12x 3.39 GHz and 46.7 GB RAM. By contrast, using the presented
305 method, an effective selection can be obtained in minutes through the following
306 steps:

- 307 • A single simulation of data, including all 20 candidate ground stations.
- 308 • An initial selection based on eigenvector embedding of the ground station
- 309 relationship network (Γ).
- 310 • A simple exhaustive search optimisation, requiring the estimation of data
- 311 transfer networks as described in Alg. 2.

312 The simple exhaustive search is described in Alg. 3 and can be summarised as follows:

Algorithm 3: Optimising selection of ground stations

Input: desired number of ground station (n_{select}), estimated data transfer network (A),
ground station relationship network (Γ)

Output: Ground station selection (Φ)

Compute communities of dynamical influence (CDI) for Γ ; // see Clark et al. (2019)

\mathbf{v}_1 = first left eigenvector of Γ

$L \leftarrow$ empty list

for each CDI community **do**

- $H \leftarrow$ set of nodes in current CDI community
- Find h such that $\mathbf{v}_1[h] = \max_{h \in H}(\mathbf{v}_1[h])$
- Insert h into L ; // add to initial selection

if $|L| \leq n_{select}$ **then**

- $\Phi \leftarrow L$
- for** $i = \text{length}(L) + 1$ **to** n_{select} **do**
- for** $g \in G$ **do**
- $\Phi_{temp} \leftarrow \Phi$
- Insert g into Φ_{temp}
- $m[g] \leftarrow$ mean consensus leadership (Eq. 4) for selection Φ_{temp}
- Find g such that $m[g] = \max_{g \in G}(m[g])$
- Insert g into Φ ; // add to initial selection

else

- Sort L in descending order according to \mathbf{v}_1 entries
- $\Phi \leftarrow L[0, \dots, n_{select}]$; // reduce initial selection

while $m < m_{prev}$ **do**

- $m \leftarrow$ mean consensus leadership (Eq. 4) for selection Φ
- $m_{prev} \leftarrow m$
- for** $i = 1$ **to** n_{select} **do**
- for** $g \in G$ **do**
- $\Phi_{temp} \leftarrow \Phi$
- Replace $\Phi_{temp}[i]$ with g
- $p \leftarrow$ mean consensus leadership (Eq. 4) for selection Φ_{temp}
- if** $p < m$ **then**
- $m \leftarrow p$
- $\Phi \leftarrow \Phi_{temp}$; // update selection

313

- 314 • Identify an initial selection from eigenvector embedding

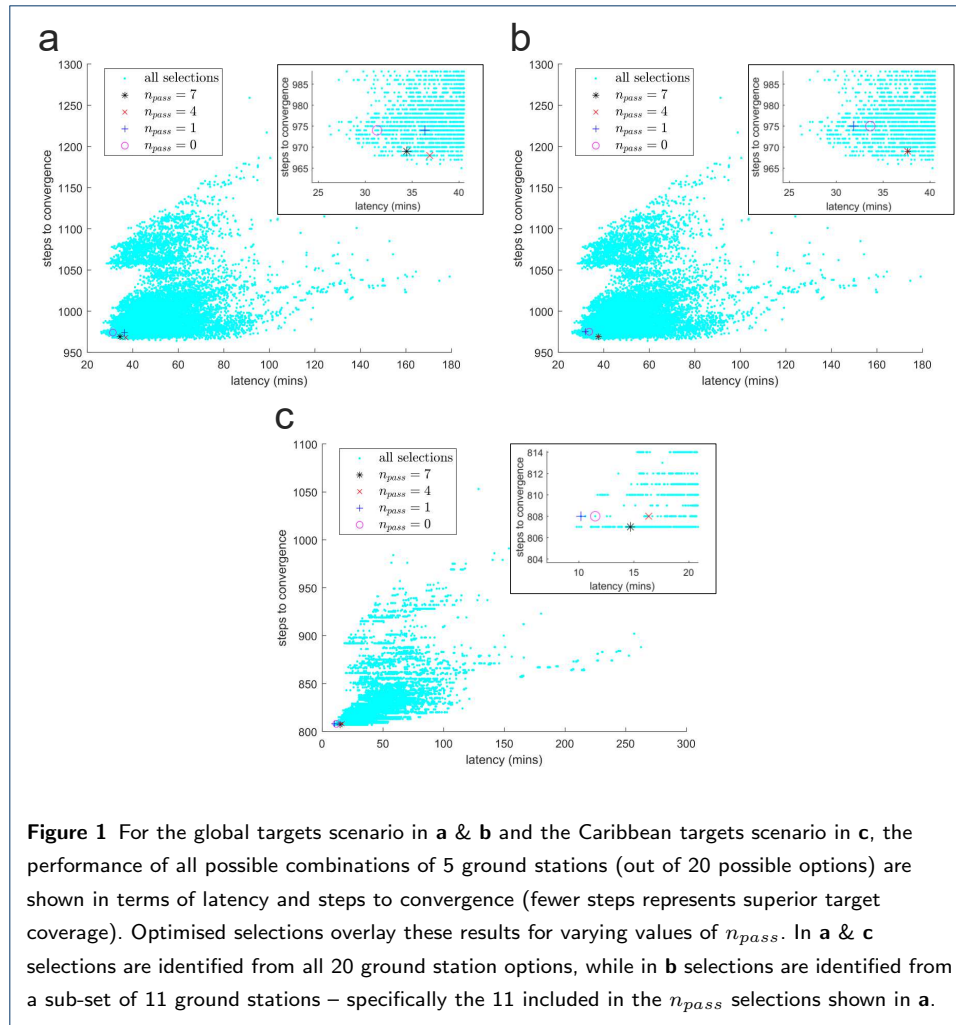
- 315 • If necessary, add to the initial selection by performing an exhaustive search
- 316 for ground stations that minimise the mean consensus leadership (Eq. 4)
- 317 • Review each selection, in turn, using an exhaustive search until the mean
- 318 consensus leadership is minimised.

319 Results

320 The efficacy of Alg. 3 is demonstrated in Fig. 1, by comparing a set of optimised
 321 selections with all possible selection combinations of five ground stations from 20
 322 possible ground station locations (geographical locations shown in Fig. 3). To as-
 323 sess the performance of selections, an individual simulation was completed for each
 324 combination detailing the movement of data over a 1-day time period to calculate
 325 the average latency (time taken from data acquisition to downlink) and the volume
 326 of data delivered from each target to each ground station. The data volumes were
 327 then used to assess the number of steps to convergence for targets under consen-
 328 sus dynamics (Eq. 2), where the connection between target and ground station is
 329 defined as being equal to the volume of data transferred. A low number of steps
 330 to convergence indicates that a ground station selection has a strong data connec-
 331 tion to all of the targets in the system (i.e. good target coverage and high data
 332 throughput).

333 In Fig. 1 the selections identified by applying Alg. 3 are seen to be near the
 334 Pareto front of the search space, with solutions producing both low latency and a
 335 low number of steps to convergence. Selections are shown for varying n_{pass} values
 336 (the number of data pass iterations, see Alg. 2). For $n_{pass} = 0$, this means that
 337 the original data transfer network is used without adaptation. The $n_{pass} = 0$ data
 338 transfer network primarily includes all of the ground stations that are the first to
 339 be seen after a satellite has collected data from a target. Note that it is possible
 340 for proceeding ground stations to also receive data from a target, but this will only
 341 occur if the spacecraft collects more data than it can downlink to the first ground
 342 station. It is therefore unsurprising that $n_{pass} = 0$ selections produce some of the
 343 lowest latency solutions.

344 In Fig. 1, the results show how $n_{pass} > 0$ can reveal selections that provide greater
 345 target coverage than $n_{pass} = 0$ selections. This is to be expected, as the estimated
 346 data transfer network will more accurately capture how data is redistributed when



347 ground stations in close proximity to targets are not selected. This allows the op-
 348 timisation to identify the ground stations that will receive the most data when
 349 reducing from 20 to 5 ground stations.

350 It can also be seen in Fig. 1 that there is variation in the results depending on
 351 the n_{pass} value. The n_{pass} value determines the number of data pass iterations
 352 when estimating the data transfer network (Alg. 2), where with each iteration data
 353 is passed on from unselected ground stations. Therefore with too few iterations
 354 insufficient data is passed on to ground stations that would form effective selections.
 355 Conversely, too many iterations results in an excess of data being estimated as
 356 arriving at poorly connected ground stations. Hence, an n_{pass} value similar to the
 357 average number of ground stations located between a target and a selected ground
 358 station is recommended. In this case, with 20 ground station locations and only 5

359 selected an $n_{pass} = 4$ would be expected to perform best for optimising steps to
360 convergence. However, as demonstrated in Fig. 1 **c** this is not a guarantee given the
361 errors in estimating data transfer and the combinatorial nature of the problem.

362 To demonstrate the performance of the method with differing initial selections,
363 Fig. 1 **b** shows the results of selecting a set of 5 ground stations from 11 candidate
364 ground stations. These 11 ground stations are a subset of the original set of 20
365 stations, selected by identifying any station that appeared in any of the $n_{pass} = 0-7$
366 selections in Fig. 1 **a**. These results show that, despite a reduced search space, the
367 $n_{pass} = 0, 1, \& 4$ selections perform slightly worse in terms of steps to convergence.
368 The pattern remains similar to Fig. 1 **a**, with the $n_{pass} = 4 \& 7$ selections performing
369 notably better in terms of steps to convergence.

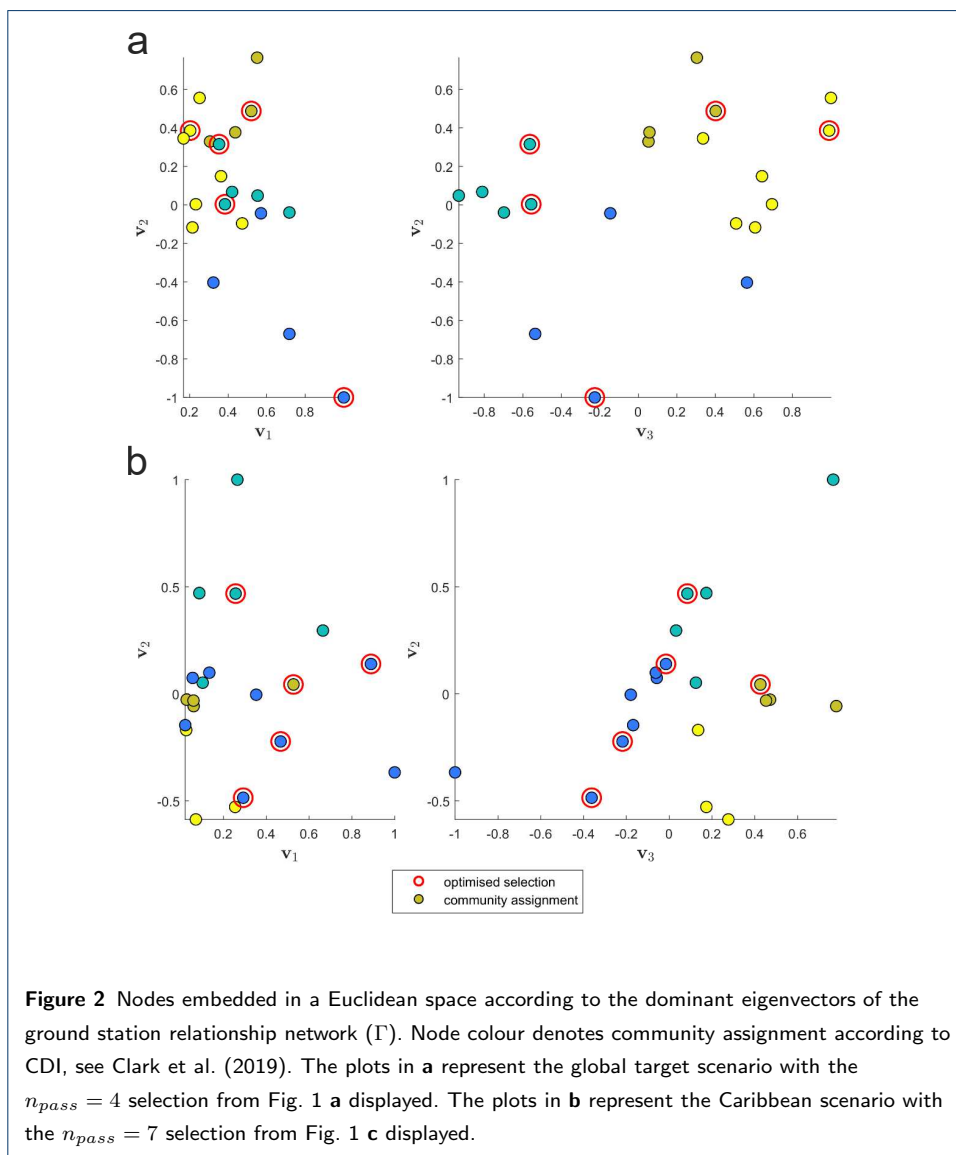
370 A similar pattern to the global targets scenario is also seen in Fig. 1 **c** for the
371 Caribbean targets scenario, whereby $n_{pass} = 0$ produces a low latency solution
372 with the lowest steps to convergence solution found by increasing n_{pass} to 7. As
373 discussed, an $n_{pass} = 4$ selection would be expected to facilitate the identification of
374 an effective selection in terms of steps to convergence. However, in this instance it is
375 likely that the localised location of targets has led to improved selection with $n_{pass} =$
376 7. Since the targets are constrained to one geographical location (Caribbean), then
377 to get an accurate picture of the data received from distant ground stations a
378 high number of iterations will be required to pass on data from unselected ground
379 stations. This is less of an issue in the global targets scenario, as most ground
380 stations are selected for their (relatively) local geographical coverage of targets.

381 Eigenvector embedding

382 The initial ground station selections, generated from eigenvector embedding of the
383 spacecraft relationship network (Γ), are altered during the optimisation to produce
384 the results shown in Fig. 1. However, Fig. 2 provides evidence that the communities
385 of dynamical influence (CDI), on which the initial selections are based, identifies
386 communities with differing target contacts that should be covered to enable good
387 target coverage. This is shown in Fig. 2, as the optimised selections cover all four
388 CDI in Fig. 2 **a** and only leave the least prominent CDI in Fig. 2 **b** unrepresented.
389 The same optimised selection can be identified by starting from a randomised ini-

390 tial selection, but eigenvector embedding-based selections reduce the number of
 391 exhaustive searches required to find an optimised solution.

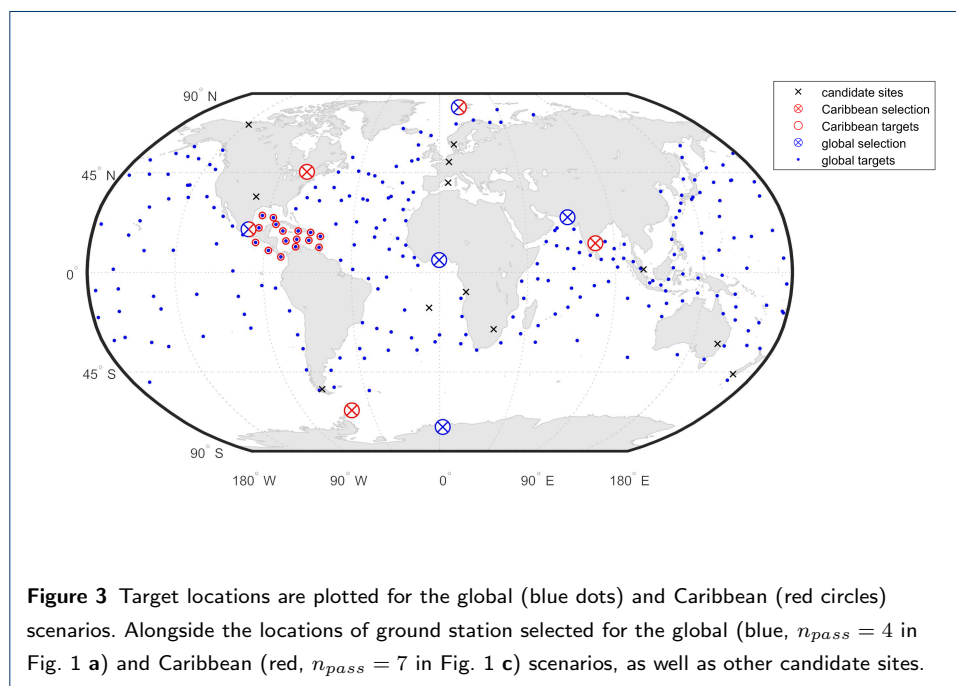
392 Ground station nodes furthest from the origin in Fig. 2 (i.e. large $\mathbf{v}_1, \mathbf{v}_2,$ & \mathbf{v}_3
 393 entries) are likely to be in receipt of large volumes of data from other ground stations
 394 in the network according to the passed-on matrices. However, the Γ matrix is only an
 395 estimation of data transfer following ground station removal. Therefore, as shown
 396 in Fig. 2b, it is possible for the node with the largest \mathbf{v}_1 entry (i.e. eigenvector
 397 centrality) to not be included in the optimised selection. This particularly occurs
 398 when other nodes in the same community are selected, as that prevents these nodes
 399 from passing data on to the most prominent node in their community.



400 Mapping the results

401 Examples of effective ground station selections are shown on a world map in Fig 3,
 402 alongside the locations of targets and all candidate ground station locations. The
 403 ground station selection for the global targets scenario (blue) forms an evenly dis-
 404 tributed cross, which facilitates the selection in achieving an even coverage of global
 405 targets through their spacecraft connections. The combination of polar and equato-
 406 rial locations is important for spacecraft connectivity, where polar ground stations
 407 achieve long connection times with the 78 polar spacecraft but cannot be relied
 408 upon exclusively as they do not receive data from the 33 other spacecraft in the
 409 constellation. The equatorial ground stations, in contrast, are seen by all spacecraft
 410 in the constellation, but for, generally, less time. This ground station selection is
 411 hence driven by the hybrid nature of the constellation.

412 The Caribbean scenario also presents a distributed cross formation, but geograph-
 413 ically localised targets alter the selection by placing two ground stations in relatively
 414 close proximity (in longitude) to North America. This is facilitated through the use
 415 of the presented approach, which captures the temporal order of connections and
 416 hence encourages the selection of stations in close proximity to the target. With
 417 the majority of spacecraft in sun-synchronous or near-polar orbits, ground stations
 418 at similar longitudes to the targets will naturally provide low-latency solutions.



419 The equatorial ground station selected in India is of value as the three equatorial
420 orbiting spacecraft will consistently overfly both the Caribbean targets and this
421 ground station. Furthermore, as its location is separated from the Caribbean tar-
422 gets by approximately 180° , all non-equatorial orbiting spacecraft that view the
423 Caribbean targets on an ascending pass, will view the equatorial ground station on
424 the descending pass, and vice versa.

425 **Conclusions**

426 This paper demonstrates that effective ground station subset selections can be iden-
427 tified, for a given space system, from a single simulation involving a set of candi-
428 date sites. Consensus dynamics provide a useful basis for optimising the selection
429 of ground stations, which can be defined as sink nodes leading a set of source nodes
430 (targets) to consensus. Comparison of how rapidly the source nodes reach consen-
431 sus provides an objective that promotes the selection of subsets with important
432 properties, namely good target coverage and high data throughput.

433 The identification of effective sink nodes from a single simulation is viable due
434 to the ability to estimate data transfer networks, for a selection of sink nodes and
435 a given set of source nodes. This estimation relies on analysis of how data is re-
436 distributed when a ground station is removed from the system. The restrictions
437 applied – to the number of times data is redistributed (passed-on) when simulating
438 the system – can prevent the optimisation from identifying globally optimal solu-
439 tions in terms of convergence to consensus. However, these restrictions are desirable
440 for sink node selection in space systems, and store-and-forward data transfer sys-
441 tems in general, as they result in the optimisation implicitly rewarding lower latency
442 connections.

443 The relationships between sink nodes, in terms of passed-on data redistribution,
444 is key to both estimating the data transfer networks and for gaining insights into
445 effective selections. Insights can be obtained into effective sink node selections,
446 through embedding-based community detection in an eigenvector-defined Euclidean
447 space. Effective selections are distributed across the detected communities, which
448 is to be expected as these communities implicitly capture ground station division
449 in terms of target coverage.

CDI - Communities of dynamical influence

450 Abbreviations

451 Availability of data and materials

452 The space system datasets analysed during the current study are available in the Zenodo repository,
453 <https://doi.org/10.5281/zenodo.5243314>.

454 Funding

455 This work was funded in part by AAC CLYDE SPACE and the European Space Agency.

456 Competing interests

457 The authors declare that they have no competing interests.

458 Author's contributions

459 RAC was involved in the conceptualisation, analysis, methodology, visualisation, and drafting of the manuscript.
460 CNM was involved in the conceptualisation, methodology, and drafting the manuscript. MM was involved in the
461 conceptualisation, drafting and revision of the manuscript. All authors read and approved the final manuscript.

462 Acknowledgements

463 Not applicable.

464 Author details

465 ¹Department of Electronic and Electrical Engineering, University of Strathclyde, George Street, Glasgow, UK.

466 ²Department of Mechanical, Aerospace, and Civil Engineering, University of Manchester, Oxford Road, Manchester,
467 UK.

468 References

- 469 Capelle, M., Huguet, M.-J., Jozefowicz, N., Olive, X.: Optimizing ground station networks for free space optical
470 communications: maximizing the data transfer. *Networks* **73**(2), 234–253 (2019)
- 471 Chen, Q., Yang, L., Liu, X., Guo, J., Wu, S., Chen, X.: Multiple gateway placement in large-scale constellation
472 networks with inter-satellite links. *International Journal of Satellite Communications and Networking* **39**(1),
473 47–64 (2021)
- 474 Clark, R.A., Macdonald, M.: Identification of effective spreaders in contact networks using dynamical influence.
475 *Applied network science* **6**(1), 1–18 (2021)
- 476 Clark, R.A., McGrath, C.N., Macdonald, M.: Dynamical influence driven space system design. In: Benito, R.M.,
477 Cherifi, C., Cherifi, H., Moro, E., Rocha, L.M., Sales-Pardo, M. (eds.) *Complex Networks & Their Applications*
478 *X*, pp. 27–38. Springer, Cham (2022)
- 479 Clark, R.A., Punzo, G., Macdonald, M.: Network communities of dynamical influence. *Scientific reports* **9**(17590)
480 (2019). doi: 10.1038/s41598-019-53942-4
- 481 del Portillo, I., Cameron, B.G., Crawley, E.F.: A technical comparison of three low earth orbit satellite constellation
482 systems to provide global broadband. *Acta Astronautica* **159**, 123–135 (2019). doi:
483 10.1016/j.actaastro.2019.03.040
- 484 Di Cairano, S., Pasini, A., Bemporad, A., Murray, R.M.: Convergence properties of dynamic agents consensus
485 networks with broken links. In: 2008 American Control Conference, pp. 1362–1367 (2008). IEEE. doi:
486 10.1109/ACC.2008.4586682
- 487 Dietrich, F.J.: The globalstar satellite cellular communication system: design and status. In: WESCON/97
488 Conference Proceedings, pp. 180–186 (1997). IEEE
- 489 Keller, H., Salzwedel, H.: Link strategy for the mobile satellite system iridium. In: Proceedings of Vehicular
490 Technology Conference-VTC, vol. 2, pp. 1220–1224 (1996). IEEE
- 491 Kim, H., Seok, Y., Choi, N., Choi, Y., Kwon, T.: Optimal multi-sink positioning and energy-efficient routing in
492 wireless sensor networks. In: International Conference on Information Networking, pp. 264–274 (2005). Springer

- 493 Lacoste, F., Guérin, A., Laurens, A., Azema, G., Periard, C., Grimal, D.: FSO ground network optimization and
494 analysis considering the influence of clouds. In: Proceedings of the 5th European Conference on Antennas and
495 Propagation (EUCAP), pp. 2746–2750 (2011). IEEE
- 496 Mazzarella, L., Lowe, C., Lowndes, D., Joshi, S.K., Greenland, S., McNeil, D., Mercury, C., Macdonald, M., Rarity,
497 J., Oi, D.K.L.: Quarc: Quantum research cubesat—a constellation for quantum communication. *Cryptography*
498 **4**(1), 7 (2020)
- 499 McGrath, C.N., Clark, R.A.: Location of ground stations, targets and spacecraft for Spire Global case study. Zenodo
500 (2021). doi:10.5281/zenodo.5243314. <https://doi.org/10.5281/zenodo.5243314>
- 501 Nath, H.N., Dhamala, T.N.: Network flow approach for locating optimal sink in evacuation planning. *International*
502 *Journal of Operations Research* **15**(4), 175–185 (2018)
- 503 Pratt, S.R., Raines, R.A., Fossa, C.E., Temple, M.A.: An operational and performance overview of the iridium low
504 earth orbit satellite system. *IEEE Communications Surveys* **2**(2), 2–10 (1999)
- 505 Safa, H., El-Hajj, W., Zoubian, H.: A robust topology control solution for the sink placement problem in wsns.
506 *Journal of Network and Computer Applications* **39**, 70–82 (2014)
- 507 Spire Global, Inc.: Spire Maritime Website. <https://maritime.spire.com/>. Accessed: 10/12/2019