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

RESEARCH

Identifying effective sink node combinations in spacecraft data transfer networks

Ruaridh A Clark[1*], Ciara N McGrath[1] and Malcolm Macdonald[1]

[*]Correspondence: ruaridh.clark@strath.ac.uk
[1]Department 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

Introduction

Historically, satellite constellations were composed of a few large spacecraft that produced simple, grid-like, communication network topologies Dietrich (1997); Keller and Salzwedel (1996); Pratt et al. (1999). In contrast, new small-satellite constellations present as complex data transfer networks due to the variety of orbital positions, as a result of a reliance on ad-hoc launch opportunities. This presents a challenge for operators to efficiently select, or locate, ground stations that can suitably service their constellation. This paper demonstrates how holistic assessment of these complex networks can provide an analytical approach to geographical ground station selection - a highly combinatorial problem. Such an approach opens up the potential for agile and responsive space systems that can be adapted by altering their connectivity to the ground, rather than relying on costly and limited spacecraft manoeuvring capabilities. While the developed approach is shown to be effective for the specific challenge of space system analysis, it is expected that the presented methods could be effectively applied to a range of similar problems seen in, for example, traffic flow systems (see Nath and Dhamala (2018)) and wireless sensor networks (see Kim et al. (2005); Safa et al. (2014)).

Data transfer is a spreading process that Clark et al. (2019) showed can be represented by a network in order to detect the relative influence of nodes. A network of averaged contacts over time enables the network's adjacency matrix to provide insights into the major pathways for spread, where Clark and Macdonald (2021) demonstrated this by identifying influential disease spreaders in contact networks. For space system flow networks, where targets are sources of data and ground stations are sinks, Clark et al. (2022) detailed how the eigenvectors of the adjacency matrix can reveal the relative influence of ground stations in terms of receiving target data. However, the aggregation of contact times, to approximate data transfer as in Clark et al. (2022), limits the applicability of the approach to a system dealing with the transfer of discrete data packets – as is the case in many applications including Earth observation and Internet of Things (IoT) services. Since the order in which a spacecraft comes into contact with targets and ground stations plays an important role in determining system performance. To address this challenge, we go beyond the work presented in Clark et al. (2022) by proposing an aggregated network that accounts for the temporal ordering of contacts. This includes the redistribution of data when a ground station is removed from the system, a necessary step in evaluating an effective subset selection from a set of candidate ground station locations. The redistribution of data provides an estimate for where data will go if a ground station is no longer included in the system, which impacts the relative influence of each sink node within the network.

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

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

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

Methods

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

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

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

Space System Scenarios

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

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

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

Generating data transfer networks

A data transfer network (Λ) is created to capture the data transactions in the space system, with a set of ground targets, spacecraft in given orbits, and a set of candidate ground stations. The network is populated by propagating the satellites' motion and simulating data transfer in the system for a defined period of time (T), during which the movement of data packets is monitored. The process of generating Λ is described in detail in Alg. 1 (in black text), and is summarised as follows:

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

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

- Each spacecraft in the system is given a second data buffer ($db2$), which is populated with dummy data (0 entries) when in contact with targets (i.e matching the data inserted into $db1$).
- In addition to dummy data, a *passed-on* data packet $[d, g]$ is inserted into $db2$ for every data packet d that is removed (i.e. downlinked) from $db1$ for the same spacecraft, where g is the current ground station (in contact with the spacecraft).
- When a spacecraft is in contact with a ground station, the dummy (0 entry) 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 Λ ← zeros(n_D + n_G, n_D + n_G)
for g ∈ G do
    Set B[g] ← zeros(n_G, n_G)
for s ∈ S do
    db1, db2 ← new data buffer lists ;                          // data buffers
    n_db1 ← 0, n_db2 ← zeros(n_G) ;                             // downlink counters
    for t ← 0 to T by τ do
        for d ∈ D do
            if d ∈ C[t] then
                Insert d into db1 ;                             // data acquired from target
                Insert 0 into db2 ;                             // dummy data
        for g ∈ G do
            if g ∈ C[t] then
                Ω ← G ∈ C[t, t + τ, ..., t + (τ × 9)] ;  // set of ground stations soon
                    to be or currently in contact with s
                while (0 < n_db1 < δ and db1 not empty) or (0 < n_db2 < δ and db2 not
                    empty) do
                    for d ∈ D do
                        if 0 < n_db1 < δ and d ∈ db1 then
                            Remove d from db1
                            Λ_{d,n_D+g} = Λ_{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_{γ∈G} (n_db2[γ]) < δ then
                            if 0 ∈ db2 then
                                Remove 0 from db2
                                for γ ∈ G do
                                    n_db2[γ] = n_db2[γ] + 1 ;   // add to downlink count
                            else
                                for γ ∈ G do
                                    if 0 < n_db2[γ] < δ and [d, γ] ∈ db2 and γ ∉ Ω then
                                        Remove [d, γ] from db2
                                        if packet re-insertion count < ρ then
                                            Insert [d, g] into db2 ;    // passed-on data
                                            Add 1 to packet re-insertion count
                                        B[γ]_{d,n_D+g} = B[γ]_{d,n_D+g} + 1 ;  // update B[γ]
                                        n_db2[γ] = n_db2[γ] + 1 ; // add to downlink count
```

packets associated with ground stations. Only once all the dummy data is removed, then the passed-on data packets are removed from $db2$.

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

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

Estimating data transfer network

The difficulty in identifying effective ground station combinations stems from the impact that one selection has on the value of other ground stations in receiving data and covering targets. For example, a ground station (GS1) may be viewed by a spacecraft that has received data from a target (T1). However, it is possible that the data transfer network (Λ) does not report this connection if, for instance, the spacecraft has already downlinked all of T1's data to other ground stations prior to overflight of GS1. Therefore, we propose an approach for estimating the data received by a subset selection of ground stations, using the passed-on networks (B) to identify where data would go if a ground station was removed. This approach has been formulated for the analysis of space systems, but such an approach is generalisable to combinatorial flow network problems, where the removal of a sink node results in greater traffic arriving at other sinks in the network.

To estimate the data received by a subset selection, the data transfer network (Λ) defined for the full set of ground stations needs to be updated according to the passed-on networks (B). This process creates an estimated data transfer network (A) and is detailed in Alg. 2, with the ground station selection represented by a vector $\mathbf{r}$ where $r_g = 1$ indicates a selected ground station g, and $r_g = 0$ denotes an unselected ground station. The process involves moving data from each unselected ground station, in turn, by using the normalised passed-on matrix (K) to determine where the data goes, before removing data from the ground station's column in the 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
        $s_d \leftarrow 1$
    $\Phi[g] = \mathrm{diag}(\mathbf{s})\, B[g]$; // row normalised
for $n \leftarrow 1$ **to** n_{pass} **do**
  for $g \in G$ **do**
    $K = \mathrm{diag}(A_{D,g})\, \Phi[g]$; // identify data movement
    $A_{D,g} = A_{D,g} - A_{D,g} \times (\max(\mathbf{r}) - r[g])$; // remove data
    $A = A + \mathrm{diag}(\max(\mathbf{r}) - r[g])\, K$; // move data
$A = \mathrm{diag}(\frac{r}{\max(r)})\, A$; // only keep data for selected ground stations

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

Consensus dynamics for target coverage

An effective way of evaluating a subset selection of ground stations in terms of target coverage and data throughput, for a network $G = (V, E)$ of targets and ground stations, is through the use of consensus dynamics. Specifically consensus leadership, where ground station selections are identified by assessing their ability to lead targets to consensus – according to the following consensus protocol – when the 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

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

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

$$x_i[t+1] = x_i[t] + \epsilon u_i[t] \tag{2}$$

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

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

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,

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

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

The optimisation can then be defined as follows,

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

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

Initial eigenvector-based selection

The optimisation of ground station selections is a highly combinatorial problem and as such susceptible to local optima far from the global optimum. This issue is exacerbated by the need to update the data transfer network for every possible selection. We propose an eigenvector embedding-based selection to act as an effective initial selection, providing an alternative to a more exhaustive search. The use of brute-force evaluation of all combinations is often intractable for sufficiently large 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, $B[g] \ \forall \ g \in G$, detail the movement of data from targets to ground station, the network Γ details the connections between ground stations, where

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

The Γ network is useful in identifying influential ground stations. This is despite Γ only detailing the relationships between ground stations, since these relationships are a product of connectivity to spacecraft that have collected target data. Therefore, Γ highlights whether ground stations are connected to spacecraft in similar or different orbits. Differing spacecraft orbits result in different target contacts, where these differences lead to different patterns of target coverage. Hence, selecting ground stations that cover different sets of spacecraft will also likely provide a selection that covers differing communities of targets.

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

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

Exhaustive search optimisation

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

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

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)
v_1 = first left eigenvector of Γ
L ← empty list
for each CDI community do
    H ← set of nodes in current CDI community
    Find h such that v_1[h] = max_{h∈H}(v_1[h])
    Insert h into L ;                                          // add to initial selection
if |L| ≤ n_select then
    Φ ← L
    for i = length(L) + 1 to n_select do
        for g ∈ G do
            Φ_temp ← Φ
            Insert g into Φ_temp
            m[g] ← mean consensus leadership (Eq. 4) for selection Φ_temp
        Find g such that m[g] = max_{g∈G}(m[g])
        Insert g into Φ ;                                      // add to initial selection
else
    Sort L in descending order according to v_1 entries
    Φ ← L[0, ..., n_select] ;                                  // reduce initial selection
while m < m_prev do
    m ← mean consensus leadership (Eq. 4) for selection Φ
    m_prev ← m
    for i = 1 to n_select do
        for g ∈ G do
            Φ_temp ← Φ
            Replace Φ_temp[i] with g
            p ← mean consensus leadership (Eq. 4) for selection Φ_temp
            if p < m then
                m ← p
                Φ ← Φ_temp ;                                   // update selection
```

- Identify an initial selection from eigenvector embedding

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

Results

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

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

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

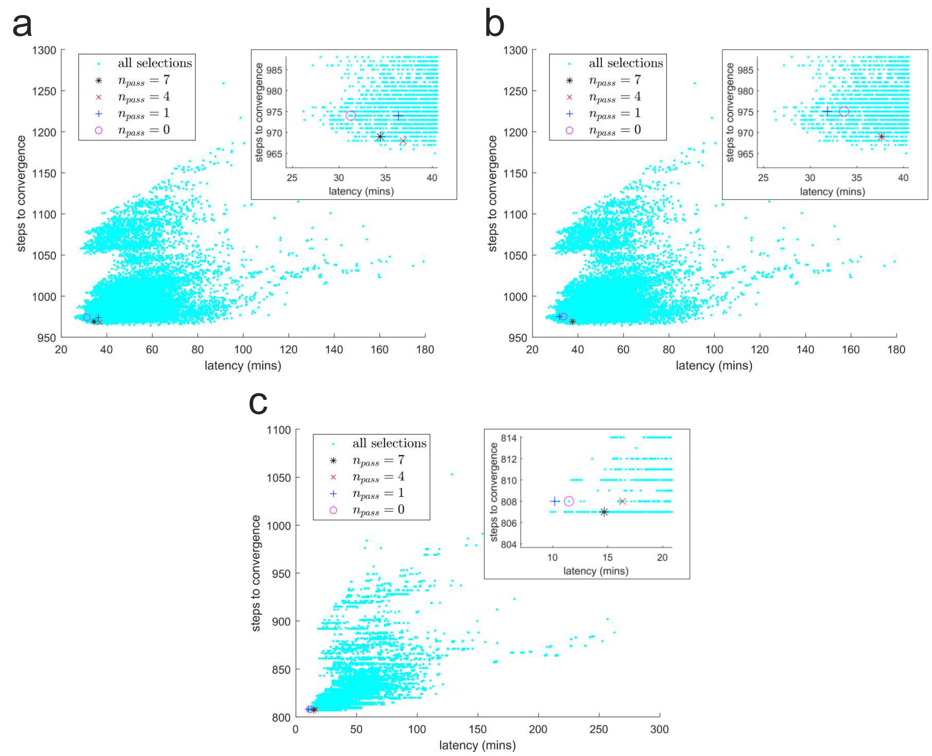

Figure 1 For the global targets scenario in **a** & **b** and the Caribbean targets scenario in **c**, the performance of all possible combinations of 5 ground stations (out of 20 possible options) are shown in terms of latency and steps to convergence (fewer steps represents superior target coverage). Optimised selections overlay these results for varying values of n_{pass}. In **a** & **c** selections are identified from all 20 ground station options, while in **b** selections are identified from a sub-set of 11 ground stations – specifically the 11 included in the n_{pass} selections shown in **a**.

ground stations in close proximity to targets are not selected. This allows the optimisation to identify the ground stations that will receive the most data when reducing from 20 to 5 ground stations.

It can also be seen in Fig. 1 that there is variation in the results depending on the n_{pass} value. The n_{pass} value determines the number of data pass iterations when estimating the data transfer network (Alg. 2), where with each iteration data is passed on from unselected ground stations. Therefore with too few iterations insufficient data is passed on to ground stations that would form effective selections. Conversely, too many iterations results in an excess of data being estimated as arriving at poorly connected ground stations. Hence, an n_{pass} value similar to the average number of ground stations located between a target and a selected ground station is recommended. In this case, with 20 ground station locations and only 5 selected an $n_{pass} = 4$ would be expected to perform best for optimising steps to convergence. However, as demonstrated in Fig. 1 **c** this is not a guarantee given the errors in estimating data transfer and the combinatorial nature of the problem.

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

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

Eigenvector embedding

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

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

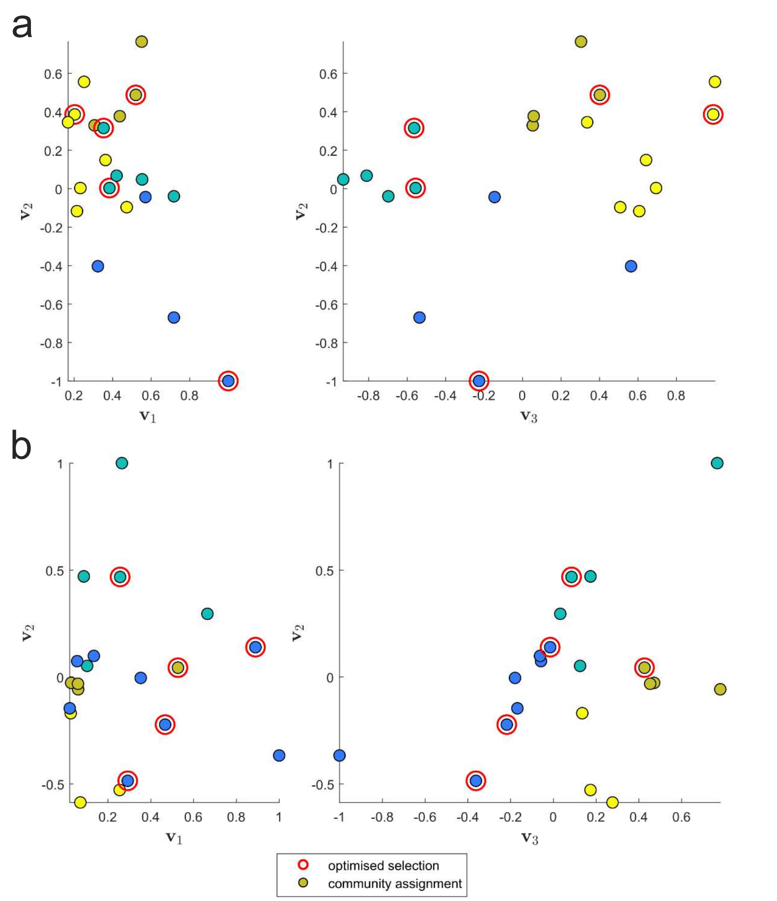

Figure 2 Nodes embedded in a Euclidean space according to the dominant eigenvectors of the ground station relationship network (Γ). Node colour denotes community assignment according to CDI, see Clark et al. (2019). The plots in **a** represent the global target scenario with the $n_{pass} = 4$ selection from Fig. 1 **a** displayed. The plots in **b** represent the Caribbean scenario with the $n_{pass} = 7$ selection from Fig. 1 **c** displayed.

Mapping the results

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

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

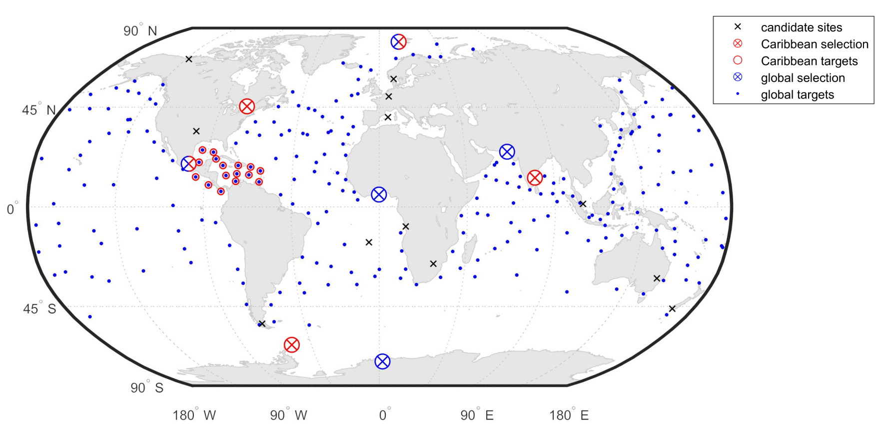

Figure 3 Target locations are plotted for the global (blue dots) and Caribbean (red circles) scenarios. Alongside the locations of ground station selected for the global (blue, $n_{pass} = 4$ in Fig. 1 **a**) and Caribbean (red, $n_{pass} = 7$ in Fig. 1 **c**) scenarios, as well as other candidate sites.

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

Conclusions

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

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

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

CDI - Communities of dynamical influence

Abbreviations

Availability of data and materials

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

Funding

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

Competing interests

The authors declare that they have no competing interests.

Author's contributions

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

Acknowledgements

Not applicable.

Author details

[1]Department of Electronic and Electrical Engineering, University of Strathclyde, George Street, Glasgow, UK.
[2]Department of Mechanical, Aerospace, and Civil Engineering, University of Manchester, Oxford Road, Manchester, UK.

References

Capelle, M., Huguet, M.-J., Jozefowiez, N., Olive, X.: Optimizing ground station networks for free space optical communications: maximizing the data transfer. Networks **73**(2), 234–253 (2019)

Chen, Q., Yang, L., Liu, X., Guo, J., Wu, S., Chen, X.: Multiple gateway placement in large-scale constellation networks with inter-satellite links. International Journal of Satellite Communications and Networking **39**(1), 47–64 (2021)

Clark, R.A., Macdonald, M.: Identification of effective spreaders in contact networks using dynamical influence. Applied network science **6**(1), 1–18 (2021)

Clark, R.A., McGrath, C.N., Macdonald, M.: Dynamical influence driven space system design. In: Benito, R.M., Cherifi, C., Cherifi, H., Moro, E., Rocha, L.M., Sales-Pardo, M. (eds.) Complex Networks & Their Applications X, pp. 27–38. Springer, Cham (2022)

Clark, R.A., Punzo, G., Macdonald, M.: Network communities of dynamical influence. Scientific reports **9**(17590) (2019). doi: 10.1038/s41598-019-53942-4

del Portillo, I., Cameron, B.G., Crawley, E.F.: A technical comparison of three low earth orbit satellite constellation systems to provide global broadband. Acta Astronautica **159**, 123–135 (2019). doi: 10.1016/j.actaastro.2019.03.040

Di Cairano, S., Pasini, A., Bemporad, A., Murray, R.M.: Convergence properties of dynamic agents consensus networks with broken links. In: 2008 American Control Conference, pp. 1362–1367 (2008). IEEE. doi: 10.1109/ACC.2008.4586682

Dietrich, F.J.: The globalstar satellite cellular communication system: design and status. In: WESCON/97 Conference Proceedings, pp. 180–186 (1997). IEEE

Keller, H., Salzwedel, H.: Link strategy for the mobile satellite system iridium. In: Proceedings of Vehicular Technology Conference-VTC, vol. 2, pp. 1220–1224 (1996). IEEE

Kim, H., Seok, Y., Choi, N., Choi, Y., Kwon, T.: Optimal multi-sink positioning and energy-efficient routing in wireless sensor networks. In: International Conference on Information Networking, pp. 264–274 (2005). Springer

Lacoste, F., Guérin, A., Laurens, A., Azema, G., Periard, C., Grimal, D.: FSO ground network optimization and analysis considering the influence of clouds. In: Proceedings of the 5th European Conference on Antennas and Propagation (EUCAP), pp. 2746–2750 (2011). IEEE

Mazzarella, L., Lowe, C., Lowndes, D., Joshi, S.K., Greenland, S., McNeil, D., Mercury, C., Macdonald, M., Rarity, J., Oi, D.K.L.: Quarc: Quantum research cubesat—a constellation for quantum communication. Cryptography **4**(1), 7 (2020)

McGrath, C.N., Clark, R.A.: Location of ground stations, targets and spacecraft for Spire Global case study. Zenodo (2021). doi:10.5281/zenodo.5243314. https://doi.org/10.5281/zenodo.5243314

Nath, H.N., Dhamala, T.N.: Network flow approach for locating optimal sink in evacuation planning. International Journal of Operations Research **15**(4), 175–185 (2018)

Pratt, S.R., Raines, R.A., Fossa, C.E., Temple, M.A.: An operational and performance overview of the iridium low earth orbit satellite system. IEEE Communications Surveys **2**(2), 2–10 (1999)

Safa, H., El-Hajj, W., Zoubian, H.: A robust topology control solution for the sink placement problem in wsns. Journal of Network and Computer Applications **39**, 70–82 (2014)

Spire Global, Inc.: Spire Maritime Website. https://maritime.spire.com/. Accessed: 10/12/2019