Optimizing classroom assignments to minimize epidemiological risk: the sibling rewiring problem

José Manuel Galán[1], Silvia Díaz-de la Fuente[2], Virginia Ahedo[1] and José Ignacio Santos[1]

[1] Departamento de Ingeniería de Organización, Universidad de Burgos, Burgos, Spain
[2] Departamento de Geografía, Área de Didáctica de las Ciencias Sociales, Universidad de Salamanca, Salamanca, Spain

ABSTRACT

In the context of infectious diseases, the assignment of students to classroom groups can significantly influence infection dynamics within school environments, particularly when sibling relationships introduce latent connections between otherwise unconnected groups. Traditional grouping methods and pandemic-era bubble strategies do not explicitly optimize student network structures or account for equity in exposure. This study introduces the Sibling Rewiring Problem, a novel multi-objective framework for student assignment that aims to maximize network fragmentation, reduce potential contagion pathways and minimize variance in group sizes and epidemiological exposure—thereby promoting fairness. We compared baseline, heuristic, and metaheuristic strategies in realistic school scenarios. A simple heuristic that assigns siblings to the same classroom line when feasible consistently achieves substantial network fragmentation with minimal impact on equity. Simulated Annealing further improved these results, particularly in complex configurations with densely connected sibling networks. Our findings suggest that family-aware classroom assignments can enhance epidemiological resilience while maintaining socially acceptable distributions. This approach provides a practical and scalable framework for integrating public health considerations into educational planning and may inform future decision-making in both emergency and routine contexts.

Subjects Algorithms and Analysis of Algorithms, Computer Education, Data Science, Network Science and Online Social Networks, Optimization Theory and Computation
Keywords Classroom assignment, Sibling networks, Epidemic risk, Assignment optimization, Network fragmentation, Multi-objective optimization, Equitable risk distribution

Submitted 4 June 2025
Accepted 28 January 2026
Published 17 March 2026

Corresponding author
José Manuel Galán, jmgalan@ubu.es

Academic editor
Jacqui Chetty

Additional Information and Declarations can be found on page 26

DOI **10.7717/peerj-cs.3710**

INTRODUCTION

The COVID-19 pandemic forced educational institutions worldwide to adopt preventive measures aimed at minimizing viral transmission while preserving the continuity of learning (*Lo Moro et al., 2020*; *Krishnaratne et al., 2020*; *Sundaram et al., 2021*; *Kaiser et al., 2021*; *Esposito, Cotugno & Principi, 2021*; *Sombetzki et al., 2021*; *Liu, Huang & Wang, 2023*; *Littlecott et al., 2024*). One of the approaches implemented in Spanish schools was the creation of stable classroom groups, commonly known as "bubble groups". The underlying idea was to significantly reduce social contact by limiting interactions to fixed groups, thereby preventing the virus from spreading across the broader student population. While initially effective, the bubble-group strategy failed to account for indirect social interactions outside the classroom setting—particularly those stemming from household contacts, such as siblings enrolled in different classrooms or grades. These overlooked latent interactions critically undermined the intended epidemiological effectiveness of the bubble strategy.

Recent reviews have emphasized the growing role of artificial intelligence, big data, and network-based modeling in epidemic intelligence and infectious disease surveillance (*Kamarul Aryffin et al., 2025*). These advances highlight the need for decision-support frameworks capable of integrating epidemiological risk with social or organizational structures such as schools.

The existence of siblings within a school creates implicit connections between seemingly independent bubble groups, forming potential pathways for virus propagation across the network of classrooms. In practical terms, even if classrooms remain separated during school hours, siblings from different groups inevitably interact at home, potentially connecting different groups through family units. To effectively address this issue, this article formulates a novel optimization problem, called the "Sibling Rewiring Problem," which seeks to strategically reassign students to classrooms by explicitly accounting for sibling relationships.

The goal of this optimization problem is twofold. First, it aims to maximize the fragmentation of the classroom interaction network by increasing the number of disconnected components, thus significantly limiting potential pathways for virus transmission. Second, our approach simultaneously seeks to minimize disparities in the sizes of these components, promoting an equitable distribution of contagion risks across the entire student body. Achieving both objectives simultaneously poses a complex multi-criteria optimization challenge that requires careful consideration and methodological rigor.

This study comprehensively explores and compares several strategies to address the sibling rewiring problem. We examined five approaches: (1) random assignment, as a baseline scenario without sibling considerations; (2) bubble strategy, assigning siblings from the same grade into the same classroom without optimizing the overall network structure; (3) a heuristic approach, assigning siblings systematically within the same classroom line to limit cross-group interactions; and two advanced metaheuristics: (4) simulated annealing and (5) a multicriteria genetic algorithm. Unlike the first three strategies, each of which yields a single solution, simulated annealing and the genetic algorithm produce sets of Pareto-optimal solutions, providing a spectrum of trade-offs between network fragmentation and equitable component size distributions.

The effectiveness of each method was systematically evaluated using 180 synthetic datasets that represent realistic scenarios in Spanish schools. These datasets vary in terms of the number of classroom lines per grade (2, 3, and 4) and simulate a broad range of family-size distributions, modeled using a Poisson distribution with a mean incrementally ranging from 0.1 to 2.0. Such detailed experimentation allows for a robust assessment of the relative performance of each strategy under different real-world conditions.

Before formally stating the problem, we clarify the type of educational contexts in which the proposed model is applicable, and provide international evidence supporting the structural assumptions of our framework. The optimization framework proposed in this study is designed for educational systems in which students are organized into stable classroom groups within each grade level. This organizational structure—whereby students remain assigned to a fixed peer group throughout the academic year—is common across many national education systems. In Spain, for instance, this is the standard arrangement in elementary and primary education: each grade is typically divided into multiple classroom units, commonly labeled A, B, C, *etc*., with each label corresponding to a distinct and stable cohort of students (*European Commission/EACEA/Eurydice, 2025*). In Spain, this subdivision is traditionally referred to as a *línea* (literally, 'line'), meaning a parallel classroom group within the same grade. Throughout this article, we use the term classroom line to denote such parallel groups, and the term group to refer generically to any classroom unit when the distinction is not essential.

While terminology and implementation details vary internationally, the underlying principle of stable, grade-specific groupings is widespread in many education systems. It is notably present in France (*European Commission/EACEA/Eurydice, 2023a*), Poland (*Herbst, Sobotka & Wójcik, 2023*), England and Wales (*Hallam, Ireson & Davies, 2004*), and Germany (*European Commission/EACEA/Eurydice, 2023b*) (from first to fourth grade in the majority of Länder), among others (see Eurydice: https://eurydice.eacea.ec.europa.eu/). In rural or low-enrollment settings, multi-grade classrooms may exist (*Hallam, Ireson & Davies, 2004*; *Casserly, Tiernan & Maguire, 2019*); however, even in these cases, group composition is relatively constrained and remains stable over time. In large schools, exceeding class size limits typically results in students being split into multiple stable groups within the same grade—reinforcing the relevance of the proposed framework. This is also the case in some schools in the United States, where each grade level often comprises multiple parallel classes, each assigned to a different teacher or teaching team, forming distinct and consistent instructional units throughout the school year (*Wellington et al., 2024*). A similar structure is observed in Japan, where each grade typically consists of several fixed classroom groups, each with its own homeroom teacher, and where class sizes are often larger than in many Western systems—commonly ranging between 30 and 40 students (*Cave, 2007*; *National Institute for Educational Policy Research (NIER), 2011*; *Ito, Nakamuro & Yamaguchi, 2020*).

Although national systems differ in terminology and administrative details, the core logic of assigning students to stable, grade-level cohorts is likely to be present in many other countries as well, albeit with local variations in class size, grouping policy, or teacher assignment models.

As such, the sibling rewiring model is broadly adaptable to a wide range of educational contexts where student cohorts are defined and maintained by grade. In contrast, the framework is not suitable for school systems with curriculum-based flexibility—such as those in which students form different peer groups for each subject—since group compositions vary continuously and are not organized around stable classroom structures.

The remainder of this article is structured as follows: 'Literature Review' provides a brief review of relevant literature and related problems; 'Formulation of the Sibling Rewiring Problem' formally defines the sibling rewiring problem and presents its mathematical and network-based formulation; 'Methodology' outlines the methodological framework, experimental setup, and optimization techniques; 'Results' presents and discusses the experimental results; 'Discussion' discusses the implications, limitations, and practical considerations; and finally, 'Conclusions and Future Work' summarizes the main findings and suggests potential avenues for future research.

LITERATURE REVIEW

The COVID-19 pandemic posed unprecedented challenges to educational systems globally, necessitating the rapid implementation of strategies to minimize infection risks while maintaining educational continuity. A significant body of literature has documented various interventions employed by schools, such as mask mandates, enhanced ventilation, social distancing protocols, and the creation of stable student cohorts known as "bubble groups" (*Danon, Lacasa & Brooks-Pollock, 2021*; *Alonso et al., 2021*; *McLeod et al., 2022*). Specifically, the bubble-group strategy gained substantial popularity due to its intuitiveness, simplicity, and practical ease of implementation, aiming to minimize interactions between students across different classrooms and, consequently, reduce potential contagion pathways (*Lessler et al., 2021*; *Alonso et al., 2021*).

However, several authors highlighted the critical limitations of bubble groups arising from indirect interactions that occur beyond classroom boundaries (*Danon, Lacasa & Brooks-Pollock, 2021*; *Fukumoto, McClean & Nakagawa, 2021*). *Danon, Lacasa & Brooks-Pollock (2021)*, for example, emphasized the significant epidemiological impact of household interactions, demonstrating how sibling relationships and household connections undermine the effectiveness of otherwise isolated bubble groups. Despite strict in-school separation, latent household networks remain a critical transmission pathway, posing a significant challenge to conventional separation strategies.

Empirical evidence regarding the efficacy of specific school interventions, such as closures or bubble strategies, is mixed. For instance, *Fukumoto, McClean & Nakagawa (2021)* found no causal effect of school closures on COVID-19 transmission rates in a rigorous causal analysis conducted in Japan, highlighting the complexities and methodological challenges in accurately assessing these interventions. This underscores the need for more robust, evidence-based methodologies to evaluate interventions in educational contexts effectively.

Network analysis provides a robust theoretical foundation for addressing epidemiological challenges of this type (*Vespignani et al., 2020*; *Herrmann & Schwartz, 2020*; *Maheshwari & Albert, 2020*; *Soriano-Paños et al., 2022*; *Burgio, Gómez & Arenas, 2023*). Modeling epidemiological problems as network optimization tasks has become increasingly common in the literature exploring network fragmentation and connectivity to control disease spread (*Pastor-Satorras et al., 2015*; *Kiss, Miller & Simon, 2017*;

Block et al., 2020). Previous studies have adopted network fragmentation as an optimization objective, either by maximizing the number of disconnected components to limit disease spread across network structures or by identifying optimal nodes for vaccination to efficiently "disconnect" the network and reduce effective contagion (*Cohen, Havlin & ben-Avraham, 2003*; *Holme, 2004*; *Gallos et al., 2007*; *Chami et al., 2017*).

Furthermore, multi-objective optimization approaches have proven effective in addressing complex real-world problems with potentially conflicting objectives. Under the umbrella of multi-objective optimization techniques, metaheuristic methods, such as simulated annealing (SA) and genetic algorithms (GA), have demonstrated significant potential for efficiently exploring large solution spaces within reasonable computation times (*Kirkpatrick, Gelatt & Vecchi, 1983*; *Deb et al., 2002*; *Talbi, 2009*). SA, which is characterized by the probabilistic acceptance of suboptimal solutions, effectively avoids local optima in complex combinatorial problems (*Kirkpatrick, Gelatt & Vecchi, 1983*; *Henderson, Jacobson & Johnson, 2003*; *Aarts, Korst & Michiels, 2005*). GA, especially multicriteria variants such as Non-dominated Sorting Genetic Algorithm II (NSGA-II) (*Deb et al., 2002*), successfully provide Pareto frontiers, explicitly revealing the trade-offs between competing optimization objectives (*Zitzler, Deb & Thiele, 2000*; *Coello Coello, 2006*).

Despite the significant and sophisticated methodological advances in network-based multi-objective optimization strategies, the educational domain has scarcely benefited from them, both in terms of improving student assignment strategies in general and mitigating the role of sibling interactions in infectious contexts specifically. The limited existing research typically relies on single-criterion heuristics or intuitive allocation rules, without explicitly addressing epidemiological risk minimization (network fragmentation) and/or equitable risk distribution (balanced group sizes). As a result, substantial gaps persist in both the theoretical formulation and empirical analysis of optimal student assignment strategies that explicitly incorporate sibling relationships.

The present study addresses these critical gaps by formally defining and empirically evaluating the sibling rewiring problem as a multi-objective network optimization task. Through a comparative analysis of diverse assignment strategies, including random assignment, traditional bubble groups, heuristic allocations, SA, and GA, this research explicitly incorporates sibling interactions, extends the existing literature, and offers comprehensive insights into optimal classroom assignment strategies under realistic educational scenarios.

FORMULATION OF THE SIBLING REWIRING PROBLEM

The sibling rewiring problem arises from the need to minimize potential cross-group interactions facilitated by sibling relationships within schools. Epidemiologically, the presence of siblings attending different classroom groups inherently creates hidden transmission pathways that connect otherwise isolated groups. Formally addressing these latent connections through strategic student group assignments becomes crucial for enhancing the effectiveness of preventive measures in educational settings.

Problem definition

The sibling rewiring problem can be formally stated as follows:

Given:

- Set $C = \{1, 2, \ldots, C\}$ of grade levels in a school.
- Each grade level $c \in C$ has G_c parallel classroom lines (groups).
- A maximum allowable number of students per classroom group, N_{max}.
- A set $S = \{s_1, s_2, \ldots, s_n\}$ of students, where each student s_i is characterized by
  - A unique student identifier: $i(s_i)$
  - A designated grade level $c(s_i) \in C$.
  - A list of siblings $S_{si} \subseteq S$ that may belong to different grade levels.

Decision Variables:

- A binary assignment variable $x_{i,g}$ where $x_{i,g} = 1$ if student s_i is assigned to group g in their corresponding grade level $c(s_i)$ and 0 otherwise.

Constraints:

1. Each student must be assigned to exactly one classroom group within the designated grade level.

$$\sum_{g=1}^{G_{c(s_i)}} x_{i,g} = 1, \quad \forall s_i \in S. \tag{1}$$

2. The number of students assigned to each classroom group must not exceed N_{max}.

$$\sum_{s_i | c(s_i) = cg(si) = g} x_{i,g} \leq N_{max,} \;\; \forall c_i \in C, \;\; \forall g \in \{1, \ldots, G_c\}. \tag{2}$$

3. Ideally, siblings within the same grade level should be assigned to the same classroom group to minimize latent cross-group interactions.

$$x_{i,g} = x_{j,g}, \quad \forall s_j \in S_{s_i \text{ such that }} c(s_j) = c(s_i), \forall g. \tag{3}$$

Network representation

To analyze and optimize classroom assignments, we model the problem as a network $N = (V, E)$, where

- Nodes V represent classroom groups, uniquely identified by grade level and group index $v = (c, g)$.
- Edges E represent potential epidemiological connections between nodes due to siblings assigned to different groups. An edge exists between nodes (c_i, g_i) and (c_j, g_j) if at least one pair of siblings is assigned to these respective groups.

Formally, edges are defined as:

$$E = ((c_i, g_i), (c_j, g_j)) | \exists s_u, s_v \in S\text{: } s_v \in S_{s_u}, c(s_u) = c_i, c(s_v) = c_j, x_{u,g_i} = 1, x_{v,g_j} = 1. \tag{4}$$

Optimization objectives

The sibling rewiring problem pursues two complementary objectives, resulting in a multi-objective optimization scenario:

Objective 1: maximizing network fragmentation

The primary objective is to maximize the fragmentation of the resulting interaction network, which is measured by the number of connected components (subgraphs):

$$\max f_1(x) = \text{number of connected components in N.} \tag{5}$$

Objective 2: minimizing component size variance (Equitable Risk Distribution)

The secondary objective is to ensure an equitable epidemiological risk distribution among students by minimizing the variance in the sizes of the resulting connected components.

$$\min f_2(x) = \text{Var}(|N_1|, |N_2|, \ldots, |N_k|) \tag{6}$$

where $|N_i|$ denotes the size (number of nodes) of the *i*-th connected component. A lower variance corresponds to more evenly-sized components, thus distributing epidemiological risk more uniformly.

Thus, the complete sibling rewiring problem can be summarized as the following multi-objective optimization task:

$\max f_1(x) = \text{number of connected components in } N$
$\min f_2(x) = \text{Var}(|N_1|, |N_2|, \ldots, |N_k|)$
Subject to: constraints Eqs. (1)–(3)

This multi-objective nature potentially produces a trade-off between fragmentation and equitable risk distribution, necessitating optimization techniques capable of exploring *Pareto-optimal* solutions.

A solution is considered *Pareto-optimal* if no other feasible configuration improves one of the objectives (network fragmentation or equity) without worsening the other. The set of all Pareto-optimal solutions forms the *Pareto front*, which represents the collection of non-dominated trade-offs between competing objectives. In scenarios with only one classroom group per grade level ($G_c = 1$), the problem loses its optimization flexibility, as no alternative configurations exist to explore.

METHODOLOGY

Experimental design

To evaluate the performance and applicability of the different optimization strategies proposed to address the sibling rewiring problem, an extensive experimental design was implemented. The experiments were conducted on a comprehensive set of synthetic datasets specifically generated to replicate the realistic conditions typically found in Spanish educational institutions.

Dataset description

A total of 180 synthetic datasets were generated to simulate the composition and distribution of students and their respective sibling relationships within a typical school environment. Each dataset represented a hypothetical school scenario with clearly defined grade levels and classroom groups. The synthetic nature of these datasets allows the systematic exploration of performance under controlled conditions while retaining the complexity and heterogeneity inherent in real-world settings.

Each dataset was structured following these characteristics:

- Grade levels: Each dataset comprised nine distinct grade levels, representative of a standard Spanish school structure: three kindergartens (early childhood education) and six primary school levels.
- Students per classroom group: Consistently fixed at a maximum of $N_{max} = 20$ students per classroom group, reflecting typical classroom sizes. While the main experiments assume equal group capacities, real schools often exhibit small variations in class size. To accommodate such heterogeneity, the formulation has been extended as described below.
- Number of classroom lines (groups): The experiments explored scenarios with two, three, and four parallel classroom groups per grade level, reflecting common organizational structures in schools.
- Family-size distribution: The number of siblings attending each school was modeled using a shifted Poisson distribution ($K = 1 + \text{Poisson}(\lambda)$), ensuring that every family included at least one child, as appropriate for school-conditional populations. The mean family-size (denoted as λ) ranged systematically from 0.1 to 2.0, in increments of 0.1, to cover a broad spectrum of sibling configurations—from predominantly single-child families ($\lambda \approx 0.1$) to dense sibling networks ($\lambda \approx 2.0$).

According to data from the Spanish National Statistics Institute (*Encuesta Continua de Hogares*) (*INE, 2020*), households with co-resident children are distributed as follows: 46.6% with one child, 44.4% with two children, and 9.1% with three or more. This corresponds to an average of approximately 1.6–1.7 children per family with children, yielding $\lambda \approx 0.6$–0.7 in the shifted Poisson model ($K = 1 + \text{Poisson}(\lambda)$). The experimental range adopted in this study ($\lambda = 0.1$–2.0) therefore encompasses realistic Spanish conditions while enabling sensitivity analyses and international comparisons. It should be noted that these figures represent national averages, and that the effective λ may vary across schools, neighborhoods, and socioeconomic contexts.

Handling heterogeneous classroom sizes

Although the datasets were initially designed with fixed classroom capacities, real-world school settings rarely exhibit perfectly homogeneous class sizes. To generalize the formulation and maintain feasibility under these conditions, we introduce dummy students—artificial individuals without siblings—so that the total number of students can be evenly distributed among available classes. These entities are used solely for balancing purposes and are removed from the final solution after optimization. This adjustment ensures that the assignment problem remains well defined even when class sizes are heterogeneous. Moreover, by adding more dummy students than the minimum required to achieve balance, the algorithm can explore a broader solution space and potentially reach higher fitness values at the cost of slightly greater variation in group size. This mechanism provides a flexible yet computationally simple way to handle heterogeneous classroom sizes before applying the heuristic and metaheuristic optimization strategies described below.

Experimental parameters

The experimental analysis was structured around three key parameters systematically varied across the datasets:

- Number of classroom groups (lines): Three distinct scenarios (two, three, and four classroom groups per grade level) were tested to reflect varying degrees of flexibility in student assignments.
- Poisson distribution mean (λ): A comprehensive exploration from low sibling-density scenarios ($\lambda = 0.1$) to high sibling-density scenarios ($\lambda = 2.0$) was conducted in increments of 0.1, resulting in 20 distinct family-size scenarios.
- Number of repetitions per configuration: To ensure the robustness of results and reduce variability caused by stochastic elements in dataset generation and algorithmic solutions, each specific configuration of lines and λ was replicated three times. Thus, $3 \times 20 \times 3 = 180$ datasets were created.

The choice of these experimental parameters was guided by both empirical evidence and practical considerations with the aim of ensuring representativeness and comprehensive coverage. Specifically, the selected number of classroom lines (2–4) and group sizes mirror actual conditions in many Spanish schools, while the Poisson distribution models variations in family-sizes, making the findings directly transferable to different real-world scenarios. Systematically varying the sibling distribution parameter (λ) from 0.1 to 2.0 guarantees comprehensive coverage of epidemiological scenarios, ranging from sparsely connected networks with minimal sibling interactions to densely interconnected clusters with higher transmission risks. Performing multiple repetitions per configuration enhances the robustness of the results, and the open availability of datasets, source code, and parameter settings ensures transparency, reproducibility, and facilitates future validation or extension by other researchers.

To facilitate the reproducibility of the experiments, all generated synthetic datasets (and the code to generate them or create new ones), the implementation of optimization algorithms, and analytical scripts are made publicly available on GitHub: https://github.com/josemagalan/Siblings-Rewiring.

Validation and robustness considerations

The baseline configuration of this study relied on a Poisson distribution to model household sizes. This choice is well established in demographic and epidemiological modeling (*Jennings, Lloyd-Smith & Ironmonger, 1999*; *Jennings & Lloyd-Smith, 2015*; *Jarosz, 2021*; *Lai et al., 2023*), as it provides a simple and tractable approximation for count data with relatively low dispersion.

To further assess the robustness of this modeling choice and to explore more heterogeneous demographic and structural scenarios, a comprehensive sensitivity analysis has been conducted and presented in Appendix 1–Robustness Analysis. This appendix extends the experimental design by testing alternative family-size distributions—empirical (EMP) based on census data obtained from *Encuesta Continua de Hogares*, (*INE, 2020*), geometric (*Johnson, Kemp & Kotz, 2005*), and negative binomial (*Casella & Berger, 2002*) occasionally used to model household sizes (*Rao et al., 1973*; *Patil & Rao, 1978*; *Al-Saleh & AL-Batainah, 2021*)—as well as variable classroom capacities (20, 30, and 40 students per class). The results confirm that the main conclusions remain stable across all scenarios, reinforcing the validity and generality of the proposed sibling rewiring framework.

Analyzed strategies

To thoroughly assess alternative solutions to the sibling rewiring problem, several student allocation strategies have been implemented and systematically evaluated. In this section, we describe each analyzed strategy in detail, beginning with two basic approaches serving as baseline scenarios: (1) Random Assignment (control) and (2) Bubble Strategy.

Random assignment (Control)

In the random assignment approach, students are allocated to classroom groups completely at random without any consideration of sibling relationships or epidemiological criteria. Each student within a grade level is assigned uniformly to one of the available classroom groups (lines), strictly adhering only to the constraint of the maximum group size. This strategy provides a neutral baseline scenario to quantify the impact of ignoring sibling relationships and epidemiological network fragmentation on student assignments.

Specifically, the random assignment process operates as follows (Algorithm 1): for each student in each grade, a random choice among the available classroom groups is made independently, ensuring that the maximum number of students per class (N_{max}) is never exceeded. This approach generates a highly interconnected interaction network, typically characterized by lower fragmentation due to multiple sibling connections bridging different groups.

Bubble strategy

This bubble strategy represents the epidemiological containment approach implemented during the COVID-19 pandemic in several Spanish schools. Under this scenario, siblings enrolled within the same grade level are assigned to the same classroom group whenever feasible, explicitly aiming to reduce cross-group interactions mediated by siblings within the same educational level, but does not address cross-grade interactions.

The process followed in the bubble strategy is straightforward (Algorithm 2): students with siblings in the same grade are grouped together, respecting the predefined maximum

Algorithm 1 Pseudocode—student random group assignment.

Input: List of students grouped by grade, number of groups per grade (G_c), maximum group size (N_{max})
Output: Assignment of students to groups

for each grade level c in C do
  Initialize empty groups $\{G_1, G_2, \ldots, G_c\}$, with G_c = number of groups available per grade
  for each student s in grade c do
    repeat
      Randomly select a group G_i from available groups
    until $|G_i| < N_{max}$
    Assign student s to group G_i
  end for
end for

Algorithm 2 Pseudocode—bubble strategy.

Input: List of students with student id, grade, list of siblings, number of groups per grade (G_c), maximum group size (N_{max}).
Output: Assignment of students to groups with siblings in the same grade placed together.

For each grade level c in C do
  Initialize empty groups $\{G_1, G_2, \ldots, G_c\}$, with G_c = number of groups available
  Identify sibling sets $\{S_1, S_2, \ldots, S_m\}$ within grade c
  For each sibling set S_j in grade c do
    Find a group G_i with sufficient space ($|G_i| + |S_j| \leq N_{max}$)
    Assign all students in sibling set S_j to group G_i
  end for
  For each remaining student s without siblings in grade c do
    repeat
      Randomly select a group G_i from available groups
    until $|G_i| < N_{max}$
    Assign student s to group G_i
  end for
end for

group capacity. Once all sibling-related assignments have been made, the remaining students (without siblings in the same grade) are distributed randomly among available groups, again adhering strictly to group size constraints. Although this method reduces direct sibling interactions across groups within a given grade (intra-grade cross-classroom interactions), it does not address cross-grade sibling interactions, nor does it explicitly optimize network fragmentation or equitable epidemiological risk distribution.

Heuristic strategy (Siblings always in the same line)

The heuristic strategy explicitly aims to minimize epidemiological interactions by consistently assigning siblings from the same family to the same classroom line across all grade levels whenever feasible. Recall that in the bubble strategy, the only intervention was to assign siblings within the same grade level to the same classroom, whereas in the heuristic strategy, interactions are addressed from the classroom line perspective across all grade levels (Algorithm 3). This systematic approach maximizes network fragmentation and reduces potential contagion pathways that can occur through sibling relationships that bridge different classroom groups and grades. However, strictly applying this heuristic can lead to practical limitations, particularly in scenarios characterized by large average

Algorithm 3 **Pseudocode—heuristic strategy.**

```
Input: List of students with student id, grade, list of siblings, number of groups per grade (G_c), maximum group size (N_max).
Output: Assignment of students to groups with siblings placed in the same classroom line.
Identify all families with more than one sibling attending the school.
Sort families by size (descending), prioritizing larger families.
for each family f in sorted list do
    Determine the set of feasible classroom lines across all sibling courses (lines with group availability in every grade attended by siblings of family f).
    if feasible lines exist then
        Randomly select one feasible line L.
        Assign all siblings from family f to classroom groups within their grades in line L.
    else
        for each sibling s in family f do
            Assign sibling s individually to any classroom line with available capacity in their grade, selecting the least occupied feasible group.
        end for
    end if
end for
```

Algorithm 4 **Pseudocode—greedy balancing for heuristic strategy.**

```
Input: Initial heuristic assignment (heuristic_assignment), maximum allowed students per group (N_max).
Output: Balanced assignment complying with maximum size constraints, preserving maximal connectivity when possible.
assignment ← heuristic_assignment
initial_components ← count_connected_components(assignment)
repeat // Phase 1: Reassign entire families without reducing connectivity
    updated ← FALSE
    Calculate Δ = n_c,g − N_max for each classroom (c, g)
    for each overpopulated classroom (Δ > 0) do
        for each family F assigned to this classroom do
            for each underpopulated classroom (Δ < 0) do
                provisional_assignment ← move_family(F, new_classroom)
provisional_components ← count_connected_components(provisional_assignment)
                if provisional_components ≥ initial_components and
                    improves_distribution(provisional_assignment) then
                    assignment ← provisional_assignment
                    initial_components ← provisional_components
                    updated ← TRUE
                    recalculate Δ for classrooms
                    break loops to reassess classrooms
                end if
            end for
            if updated then break
        end for
        if updated then break
    end for
until not updated or no overpopulated classrooms remain
repeat // Phase 2a: Reassign individual students without reducing connectivity
    updated ← FALSE
    Calculate Δ = n_c,g − N_max for each classroom (c, g)
    for each overpopulated classroom (Δ > 0) do
        for each student s in the overpopulated classroom do
            for each underpopulated classroom (Δ < 0) do
```

Algorithm 4 (continued)

```
                provisional_assignment ← move_student(S, new_classroom)
provisional_components ← count_connected_components(provisional_assignment)
                if provisional_components ≥ initial_components and
                  improves_distribution(provisional_assignment) then
                  assignment ← provisional_assignment
                  initial_components ← provisional_components
                  updated ← TRUE
                  recalculate Δ for classrooms
                  break loops to reassess classrooms
              end if
          end for
          if updated then break
      end for
      if updated then break
  end for
until not updated or no overpopulated classrooms remain
if all classrooms are correctly balanced (Δ = 0) then
    return assignment
end if
repeat // Phase 2b: Reassign individual students allowing connectivity reduction
    updated ← FALSE
    Calculate Δ = n_{c,g} − N_max for each classroom (c, g)
    for each overpopulated classroom (Δ > 0) do
        for each student S in the overpopulated classroom do
            for each underpopulated classroom (Δ < 0) do
                provisional_assignment ← move_student(S, new_classroom)
                if improves_distribution(provisional_assignment) then
                  assignment ← provisional_assignment
                  updated ← TRUE
                  recalculate Δ for classrooms
                  break loops to reassess classrooms
                end if
            end for
            if updated then break
    end for
    if updated then break
  end for
until not updated or no overpopulated classrooms remain
Note: improves_distribution = (Σ|Δ_provisional| < Σ|Δ_actual|)
```

family-sizes (high values of the Poisson parameter λ). Specifically, grouping entire families consistently in the same line can result in certain classrooms exceeding the maximum allowed number of students (*Galán et al., 2026*).

To address the issue of overpopulated classrooms, a greedy balancing procedure was implemented to redistribute students. This balancing step involves sequentially reassigning families or, if necessary, individual students from overcrowded groups to groups with available capacity, prioritizing adjustments that preserve or minimally reduce network fragmentation. The heuristic first attempts to reassign entire families without reducing the number of connected components. If a balanced configuration cannot be achieved through family reassignment alone, the heuristic then allows the reassignment of individual students, initially favoring fragmentation-preserving moves and ultimately allowing minor fragmentation loss to meet classroom size constraints (Algorithm 4).

Additionally, an alternative heuristic variant was explored: systematically grouping all students with siblings into the same line whenever possible. While this variant significantly increases network fragmentation, thus effectively minimizing contagion pathways, it also introduces notable social equity concerns. Students belonging to families with siblings face disproportionately higher infection risks, grouped densely within the same lines, whereas students without siblings benefit from substantially lower risk exposure. Due to this disparity, such an approach might be considered socially unfair (*Fehr & Schmidt, 1999*; *Marmot, 2005*; *Tricomi et al., 2010*). Nonetheless, this extreme fragmentation strategy was strategically included in the generation of initial solutions for the subsequently implemented NSGA-II algorithm, ensuring a comprehensive exploration of the solution space and facilitating the identification of balanced trade-offs between network fragmentation and equitable risk distribution.

Simulated annealing (SA) optimization

Simulated annealing (*Kirkpatrick, Gelatt & Vecchi, 1983*) is a stochastic optimization algorithm inspired by the physical process of annealing in metallurgy. It explores the solution space by occasionally accepting inferior solutions during early iterations to escape local optima, progressively reducing this acceptance probability as the algorithm cools. This probabilistic mechanism enables broader exploration of the search landscape and helps avoid premature convergence to suboptimal configurations.

In this approach, the solution obtained from the greedy heuristic balancing strategy is used as the starting point for further optimization using the Simulated Annealing (SA) metaheuristic. The primary goal of this technique is to explore the solution space more extensively, potentially escaping local optima by probabilistically accepting worse solutions at intermediate steps, particularly in the early stages of the optimization process.

The implemented SA algorithm follows a well-defined structure. First, a neighborhood operator generates new candidate solutions by randomly selecting two students from different groups within the same grade level and swapping their group assignments. This neighbor-generation process guarantees local changes that preserve feasibility with respect to the maximum group size constraints.

The fitness function guiding the SA algorithm integrates two optimization criteria in a weighted manner: (1) maximizing the number of connected components to enhance epidemiological fragmentation and (2) minimizing the variance of component sizes to promote equitable risk distribution. Although SA is traditionally designed to optimize a single objective function, a multi-objective approach was adopted in this study by performing a weighted sweep of the two criteria. The algorithm iteratively explores a series of weighted combinations of both objectives, systematically adjusting their relative importance using weights ranging from 0 to 1 in increments of 0.1 (*e.g.*, 0.0, 0.1, 0.2, …, 1.0). Each weight combination yields a distinct Pareto-optimal candidate solution, ultimately providing a diverse set of solutions that illustrate explicit trade-offs between the two conflicting criteria.

Regarding the internal parameters of the implemented algorithm, the initial temperature (T_0) was set proportionally to the initial fitness ($T_0 = 0.3 \times$ initial_fitness), ensuring an adaptive and meaningful cooling schedule. The final temperature (T_f) was defined as 0.01, and the maximum number of iterations was set to 10,000 for a thorough exploration of the solution landscape.

To further enhance solution quality, a local search algorithm based on a first-improvement strategy was applied as a postprocessing step. This algorithm evaluates potential local improvements by sequentially evaluating a subset of possible student pair swaps until it identifies a move that improves overall fitness. Incorporating this final local search step guarantees that each solution returned by SA optimization is at least a local optimum, thereby significantly refining the obtained solutions.

Multi-objective genetic algorithm (GA) (NSGA-II)

Given the specific characteristics of the sibling rewiring problem, a customized implementation of the NSGA-II multi-objective genetic algorithm was developed (*Deb et al., 2002*). Traditional genetic algorithms rely heavily on continuous or easily permutable discrete encodings, efficient crossover mechanisms, and straightforward initialization processes. However, the sibling rewiring problem poses significant challenges that hinder standard genetic formulations, such as the discrete nature of student group assignments, the complex interactions induced by sibling relationships, and stringent classroom size constraints.

Initially, a permutation-based encoding was explored, representing solutions as vectors in which each position corresponded to a student within a grade, and the value indicated the assigned classroom group. Under this encoding, a complete solution involved a set of vectors, one per grade. However, this representation leads to overly complex solution spaces and destructive permutation-based crossover operations, resulting in poor convergence and suboptimal outcomes. Consequently, a more structured and customized genetic approach was adopted, which is detailed as follows.

In this revised implementation, each solution (individual) is represented directly as a data structure containing explicit classroom assignments for each student, organized by family. The initial population was carefully generated rather than randomly assigned. Specifically, half of the initial solutions were produced using an assignment based on a greedy balancing heuristic similar to the one described earlier (see Algorithm 4), ensuring feasible and reasonably optimized starting points. The remaining half of the initial population consists of deliberately fragmented solutions, attempting to maximize the number of disconnected components by consistently placing as many sibling groups as possible in the same classroom lines, albeit potentially creating socially imbalanced solutions.

Because of the discrete and family-oriented nature of the problem, genetic operators have been tailored accordingly. Specifically, the crossover operator acts primarily at the family level; each family is inherited from one of the two parent solutions according to a given probability parameter (prob_familia). This mechanism ensures meaningful recombination while preserving family integrity within classroom assignments; that is, siblings assignments are consistently maintained within the same line. After crossover, a capacity-based repair step guarantees compliance with the maximum allowed classroom size by relocating students from overcrowded groups to less populated alternatives within the same grade level.

The mutation operator is implemented as a simple swap mutation, randomly exchanging the classroom assignments of pairs of students within the same grade. This approach maintains feasibility while introducing variability into the solutions.

The main parameters utilized in this customized NSGA-II implementation are as follows:

- **Population size** (pop_size)**:** 50.
- **Number of generations** (num_generations)**:** 1,000.
- **Mutation probability** (p_mut)**:** 0.1 (probability of swapping classroom assignments).
- **Family-level crossover probability** (p_familia)**:** 0.5 (probability of inheriting family assignments from one parent during crossover).
- **Crossover probability** (p_cross)**:** 0.2 (probability of applying crossover between two parent solutions).

Each generation undergoes non-dominated sorting and crowding-distance selection, a characteristic of the NSGA-II algorithm, to systematically identify and preserve Pareto-optimal solutions that balance the two conflicting objectives: maximizing the number of connected components (network fragmentation) and minimizing the variance of component sizes (equitable risk distribution).

A detailed pseudocode summarizing the customized NSGA-II implementation is presented below, outlining the initialization, customized genetic operators, and Pareto-based selection process (Algorithm 5).

RESULTS

This section presents the experimental results obtained by applying the different classroom assignment strategies discussed above to a diverse set of synthetic school configurations. The strategies were evaluated using two primary optimization criteria: (1) maximizing the number of connected components in the classroom interaction network—as a proxy for epidemiological fragmentation; and (2) minimizing the variance of component sizes—to ensure equitable risk distribution among students. Additionally, the contribution of each strategy to the set of Pareto-optimal solutions was analyzed to assess their overall efficiency in the multi-objective context.

Fragmentation of the network

Figure 1 illustrates a qualitative comparison of the resulting networks under three assignment strategies for a representative scenario (λ = 0.2, 20 students per class, four classroom lines). The random initial assignment yielded a single highly connected component, indicating minimal fragmentation. The bubble strategy, although intended to reduce cross-group sibling interactions, also resulted in a single component. In contrast,

Algorithm 5 Pseudocode—customized NSGA-II.

Input: Number of generations (num_generations), Population size (pop_size), Mutation probability (p_mut), Family crossover probability (p_familia), Crossover probability (p_cross)
Output: Set of Pareto-optimal solutions balancing network fragmentation and equitable component sizes.
1: INITIALIZE population P of size pop_size:
2: - Half generated via heuristic balancing assignment
3: - Half generated maximizing sibling fragmentation
4: Evaluate fitness of each solution in population P
(number of connected components, variance of component sizes)
5: **for** generation = 1 **to** num_generations do
6: Offspring population Q ← Ø
Crossover operator (family-level):
7: **while** |Q| < pop_size do
8: **Select** two parent solutions (P_1, P_2) randomly from P
9: With probability p_cross perform crossover:
10: **for each** family f do
11: With probability p_familia assign family f according to the assignment in P_1, otherwise assign f following the assignment scheme from P_2
12: **end for**
13: Perform capacity-based repair on offspring solution
14: Add offspring solution to Q
15: **end while**
Mutation operator:
16: **for each** solution s in Q do
17: With probability p_mut, randomly select two students in the same course and swap their classroom assignments (maintaining feasibility)
18: **end for**
19: Combine populations R = P ∪ Q
20: Perform non-dominated sorting on population R into Pareto fronts
21: Calculate crowding distance for solutions in each front
22: Select next generation P by:
23: Adding solutions from Pareto fronts sequentially, starting from the best, **until** pop_size solutions are selected
24: If the last front to be added cannot fully fit into P, select solutions from that front based on descending crowding distance until pop_size is reached
25: **end for**
26: **Return** solutions from the first Pareto front as final set of solutions

the heuristic strategy with greedy balancing produced four disconnected components, demonstrating greater effectiveness in fragmenting the network.

A deeper analysis of network fragmentation across all scenarios is presented in Fig. 2. The heuristic strategy consistently achieves a higher number of components across all values of λ and group configurations. The bubble strategy only shows a clear advantage over the initial random assignment in configurations with four lines and low λ values. In most other cases, both strategies (random and bubble group) perform similarly. A Friedman test confirmed the statistical significance of these differences ($\chi^2 = 324.73$, df = 2, $p < 2.2e{-}16$), thus validating the superiority of the heuristic approach in terms of fragmentation.

Equity in risk distribution

Although the heuristic strategy increases fragmentation, it also leads to a higher variance in the sizes of the connected components, as illustrated in Fig. 3. This effect was more

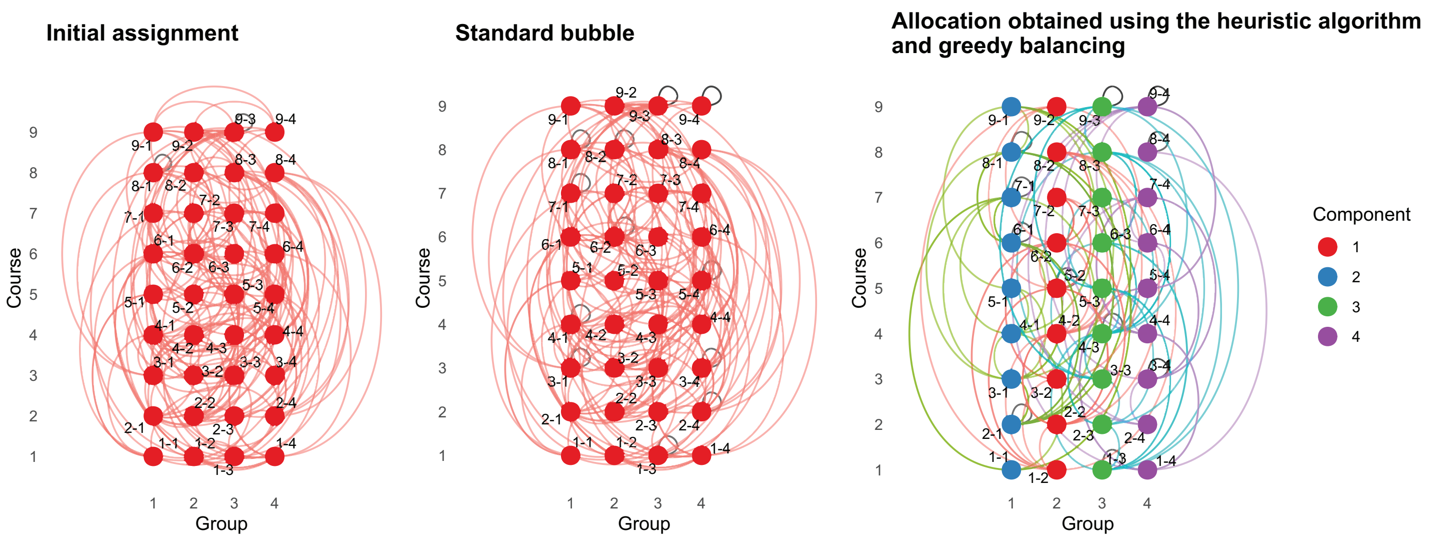

Figure 1 Comparison of classroom assignment strategies for a scenario with Poisson parameter (λ = 0.2), 20 students per group, and four parallel classroom lines. The left plot shows the initial random assignment (control), resulting in a single interconnected component (highlighted in red). The center plot represents the standard bubble strategy, shown to still maintain a single connected component. The right plot illustrates the heuristic strategy with greedy balancing, achieving significant fragmentation with four distinct connected components, improving epidemiological containment compared to previous strategies. Full-size DOI: 10.7717/peerj-cs.3710/fig-1

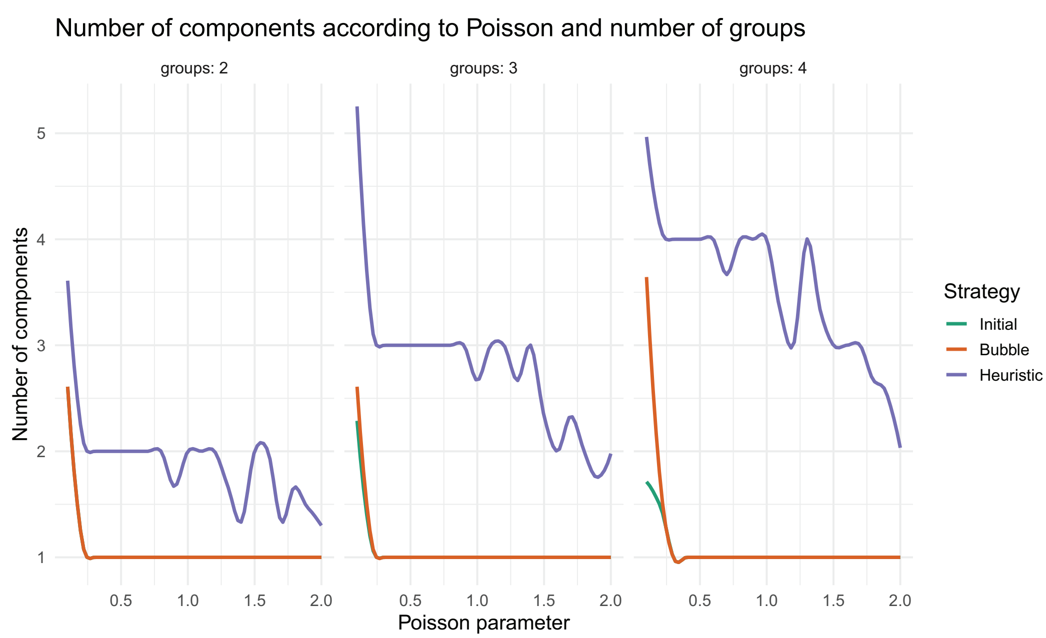

Figure 2 Number of connected components obtained for different classroom assignment strategies (Random assignment, Bubble strategy, and Heuristic with greedy balancing) across varying values of the Poisson parameter (λ) and different numbers of classroom group. The heuristic strategy consistently achieves greater network fragmentation across all scenarios. The bubble strategy only outperforms the initial assignment in cases with four classroom lines and very low values of λ; in most other cases, its performance in terms of fragmentation is nearly equivalent to that of the random assignment. Full-size DOI: 10.7717/peerj-cs.3710/fig-2

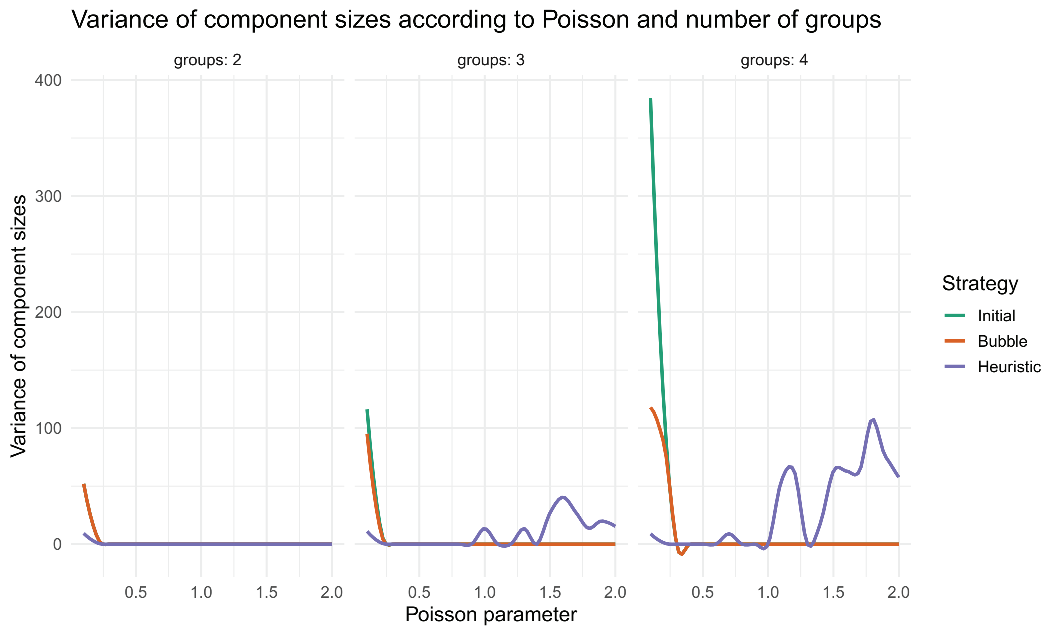

Figure 3 Variance of component sizes across different classroom assignment strategies (Initial assignment, Bubble strategy, and Heuristic with greedy balancing), as a function of the Poisson parameter (λ) and the number of classroom groups (2, 3, and 4). The heuristic strategy, by successfully fragmenting the network (as shown in the previous figure), tends to produce higher variance in component sizes, especially for larger λ values. In contrast, both the initial and bubble strategies yield more balanced component sizes (lower variance), but they do so at the cost of maintaining a highly connected network—exposing all students to similarly high levels of epidemiological risk. Notably, for scenarios with four classroom lines and low λ values, the bubble strategy outperforms the initial assignment by achieving both a higher number of components and lower variance.

Full-size DOI: 10.7717/peerj-cs.3710/fig-3

pronounced for larger values of λ, where the presence of large sibling groups imposed rigid constraints on the assignment process. On the other hand, both the initial and bubble strategies yield more uniform component sizes (lower variance), but this comes at the cost of maintaining a highly connected network in which all students face an equally high risk of contagion. Notably, the bubble strategy dominated the initial assignment only in the case of four lines and low λ values, achieving both higher fragmentation and lower variance.

Pareto-optimality and strategy comparison

The number of distinct Pareto-optimal solutions is relatively small, with many of them being structurally different yet equivalent in terms of objective values. This opens the door to incorporating additional criteria (*e.g.*, academic performance, logistical simplicity) without sacrificing solution quality under the formal objectives.

Figure 4 presents the average number of Pareto-optimal points contributed by each strategy. Metaheuristic approaches clearly dominate: the GA contributes more points in low-λ settings, primarily due to its initialization strategy that favors maximum fragmentation, whereas SA performs better in high-λ and multi-line scenarios, benefiting

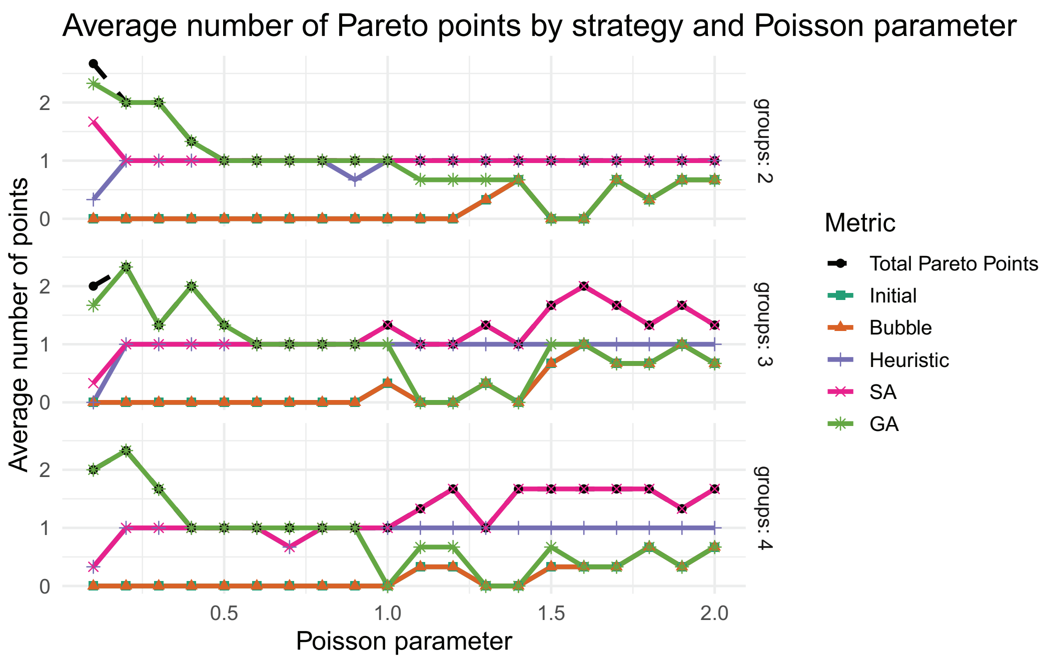

Figure 4 Average number of Pareto-optimal points contributed by each strategy (Initial, Bubble, Heuristic, Simulated Annealing (SA), and Genetic Algorithm (GA)) across different values of the Poisson parameter (λ) and classroom configurations (2, 3, and 4) line. Higher values indicate better performance, as they correspond to a greater number of non-dominated solutions contributing to the Pareto front—that is, solutions that improve one objective (fragmentation or equity) without worsening the other. Metaheuristic approaches (SA and GA) consistently dominate the Pareto fronts. GA performs particularly well in low-λ scenarios—where sibling connections are sparse—largely due to its initialization strategy that favors network fragmentation. In contrast, SA contributes more Pareto-optimal solutions in higher-λ and multi-line scenarios, benefiting from its ability to refine solutions through local exploration. Baseline strategies contribute minimally, while the heuristic approach often yields balanced, practically useful solutions despite not being explicitly optimized. Full-size DOI: 10.7717/peerj-cs.3710/fig-4

from enhanced local exploration. Baseline strategies contribute minimally to the Pareto front, with the heuristic method occasionally producing satisfactory nondominated solutions.

Figure 5 complements this analysis by showing the percentage of instances in which each strategy contributes at least one Pareto-optimal solution. As expected, GA and SA dominate the fronts, with GA being especially effective in sparse sibling scenarios, and SA gaining ground as complexity increases. Despite its simplicity, the heuristic strategy regularly yields solutions that belong to the Pareto front in less complex cases. In contrast, the initial and bubble strategies rarely (almost never) yield competitive solutions with respect to the formalized objectives.

Overall, these results demonstrate that, under realistic Spanish classroom and household conditions—typically 20–25 students per class and an average of 1.6–1.7 children per family (*INE, 2020*)—the proposed heuristic and metaheuristic strategies achieve a substantial structural reduction in potential contact pathways. The observed fragmentation levels indicate that epidemiological risk could be significantly contained

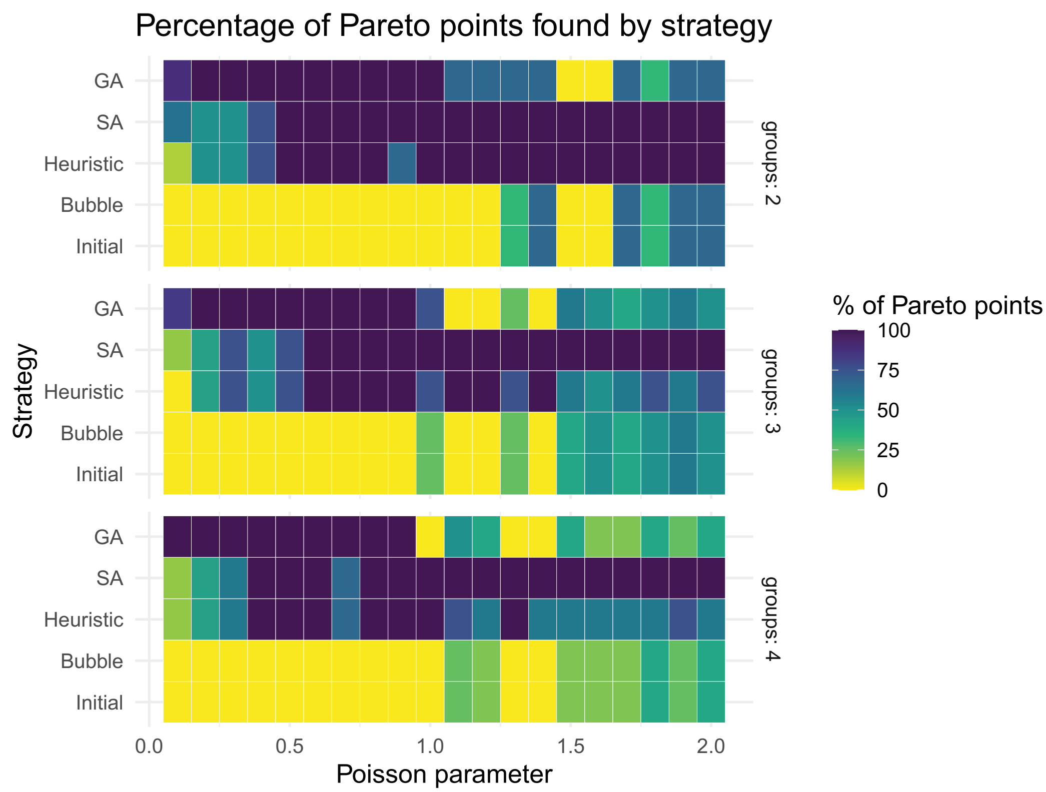

Figure 5 Percentage of Pareto-optimal points identified by each strategy (Initial, Bubble, Heuristic, Simulated Annealing (SA), and Genetic Algorithm (GA)) across varying Poisson parameters (λ) and different numbers of classroom lines (2, 3, and 4). Higher percentages indicate stronger and more consistent performance across scenarios. Metaheuristic approaches (GA and SA) consistently outperform baseline strategies, particularly as λ increases and sibling connections become denser, making the problem more complex. GA performs best in low-λ conditions with simpler family structures, whereas SA becomes increasingly effective in high-density contexts. The heuristic approach, although simpler, frequently yields at least one Pareto-optimal solution in less demanding configurations, while Initial and Bubble strategies seldom contribute to the Pareto front. Full-size DOI: 10.7717/peerj-cs.3710/fig-5

within smaller, largely independent classroom subnetworks, while preserving feasible and balanced class sizes consistent with actual school organization.

DISCUSSION

The results presented in the previous section offer a comprehensive understanding of the performance of various classroom assignment strategies across different family structures and school configurations. The findings confirm that both heuristic and metaheuristic approaches significantly outperform simpler baseline strategies in terms of network fragmentation, which is a key proxy for reducing epidemiological risk through structural separation.

Although the datasets and parameters used in this study were inspired by typical Spanish school configurations, the sibling rewiring framework is general by design. All input variables—including the number of classroom lines, maximum classroom size, and family-size distribution—can be freely adjusted to represent educational systems with different demographic, organizational, or cultural characteristics.

This flexibility allows the model to simulate conditions found in countries with larger average household sizes, varying student–teacher ratios, or alternative classroom organization policies. Consequently, the approach is not specific to Spain but can be transferred to a wide range of international contexts through straightforward reparameterization of input data.

Future research will focus on cross-country calibration using open demographic datasets (*e.g*., census-based household composition statistics) to evaluate the model's robustness and generalizability across diverse socioeconomic settings.

From an interpretative perspective, the heuristic strategy stands out as a practical and computationally efficient option. Its design, which prioritizes placing siblings within the same classroom lines, proves highly effective in fragmenting the network. However, this improvement comes at the cost of a slight reduction in equity, as it tends to produce more variance in the sizes of connected components compared with baseline approaches. Nonetheless, in most scenarios, all students are exposed to a lower overall risk than in the baseline strategies, making the minor equity loss acceptable and justifiable. This trade-off remains reasonable and ethically sound, especially when contrasted with alternative strategies that enforce strict grouping of all sibling families into the same line. Although the latter achieves even greater fragmentation, it does so by disproportionately concentrating epidemiological risk among students with siblings, leading to a significant and arguably unacceptable loss of equity.

The GA yielded promising results, particularly in scenarios with sparse sibling relationships (low λ values). However, this may be attributable to the heuristics used to generate the initial population—specifically designed to maximize fragmentation—rather than to the evolutionary process itself. Recall that the encoding schemes used in this study struggled to handle the high level of interdependencies inherent in the problem, and crossover operations often failed to preserve critical relational structures. As a result, alternative crossover configurations had to be implemented, such as copying the entire student assignment solution from one parent with probability p_familia. Although alternative encodings may offer better performance, the formulations explored here proved inefficient for this highly constrained and relationally dense problem. This limitation became especially apparent in more complex instances with high sibling density, where maintaining feasibility during crossover required extensive and costly repair procedures to respect group-size constraints.

The SA strategy struck an effective balance between exploration and exploitation, yielding competitive solutions, particularly in high-λ settings. Although SA is inherently designed for single-objective optimization, a weighted objective sweep was implemented to adapt it to the multi-objective nature of this problem. Starting from the heuristic solution, the SA algorithm was able to slightly refine the results, especially in complex configurations. Moreover, when equity is not a strict requirement, alternative initial solutions—such as assigning all siblings to the same line—can be used to explore additional trade-offs. The strong performance of SA in this context suggests that trajectory-based metaheuristics may offer particularly promising avenues for solving the sibling rewiring problem.

All strategies were evaluated under realistic conditions inspired by the Spanish educational system, including typical classroom sizes, number of lines per grade, and plausible distributions of family-sizes modeled using a Poisson process. This variety of configurations enhances the external validity of the results. However, note that the current model does not incorporate behavioral, psychological, or institutional factors that could influence real-world implementation.

The sibling rewiring framework is specifically designed for the educational stages in which bubble groups were officially implemented in Spain—pre-primary and primary education. In these stages, it is highly common for siblings to attend the same school, both for logistical reasons and because admission criteria explicitly prioritize students who already have siblings enrolled in the same center. Therefore, the assumption that siblings share the same school accurately reflects the real structure of this educational context. Higher educational levels were not organized in bubble groups but instead followed general public-health measures such as mask mandates, ventilation, and distancing. The model thus focuses on the only population segment where individuals effectively belonged to two overlapping bubble networks (family and classroom).

From a practical and ethical standpoint, the deployment of these strategies raises broader concerns beyond epidemiological optimization. Educational continuity, parental preferences, and perceived fairness must all be considered. For instance, some families may prefer to separate siblings into different classrooms for pedagogical or personal reasons, thereby introducing a potential trade-off between individual preferences and collective health. Conversely, grouping all siblings together may unfairly concentrate risk. Furthermore, changes in classroom assignments could disrupt academic continuity, especially if students are reassigned to new teachers or unfamiliar peer groups (*Wentzel, 1999*; *Rockoff, 2004*; *Benner, 2011*). Such disruptions may impose emotional, cognitive, or logistical burdens, particularly on younger children. These factors, which may isolate students or erode existing support structures, should be evaluated carefully.

Beyond methodological performance, the ethical dimension of classroom reassignment therefore deserves particular attention. Reallocating students—even when guided by epidemiological optimization—may alter established social ties and affect emotional well-being, reinforcing the need for any intervention to balance health objectives with pedagogical continuity. The sibling rewiring framework should thus be understood as a decision-support tool to help policymakers and educators evaluate the potential benefits and trade-offs of alternative grouping strategies. The optimization model is flexible and can incorporate additional objectives or constraints to mitigate such effects—for example, preserving partial social structures when friendship or peer networks are known, maintaining group stability according to academic performance, or accounting for other pedagogical and socio-emotional factors. Any practical implementation must involve educational stakeholders, prioritize student welfare, and remain consistent with pedagogical and psychological principles. Ultimately, the framework's value lies in supporting informed, context-sensitive decision-making rather than prescribing automatic or disruptive reassignments.

CONCLUSIONS AND FUTURE WORK

Main conclusions

This study introduces the Sibling Rewiring Problem, a novel formulation that explicitly incorporates sibling relationships into the optimization of classroom assignments, aimed at minimizing epidemiological risk in school settings. The problem is defined as a multi-objective optimization task with two potentially conflicting goals: (1) maximizing network fragmentation to reduce cross-group transmission potential, and (2) minimizing the variance in component sizes to ensure an equitable distribution of risk among students.

A comprehensive empirical evaluation was carried out using 180 synthetic datasets specifically designed to reflect realistic school environments, with variations in the number of classroom groups and the distribution of sibling family size. The performances of both heuristic and metaheuristic approaches, including the customized implementation of NSGA-II, were systematically assessed. The key findings may be summarized as follows:

- **Baseline strategies** such as random assignment and traditional bubble grouping offer limited epidemiological protection. Their inability to effectively fragment the student interaction network results in uniformly high exposure for all students.
- The **heuristic strategy** consistently achieves high levels of fragmentation at a very low computational cost. It provides a simple, transparent, and easily implementable approach that generally delivers improved outcomes for all students. Although it may introduce a slight increase in component size variance, the overall reduction in exposure risk makes it a highly attractive and practical solution for real-world scenarios.
- The **GA** performs well in low-density sibling scenarios; however, its success is primarily attributed to the diversity of its initial populations rather than evolutionary improvement through generations. The high degree of interdependency and structural complexity in the problem limits the effectiveness of conventional genetic operators, suggesting that the GA may not be the most suitable optimization method in this context.
- **SA**, particularly when enhanced by local search refinement, demonstrates a strong capacity to improve heuristic solutions. Its ability to effectively explore solution neighborhoods makes it especially promising for complex configurations. Moreover, it opens the possibility of starting from either balanced heuristic solutions or high-fragmentation configurations depending on the desired coverage of the Pareto front.
- The extended robustness analysis presented in Appendix 1 confirms that the framework's conclusions remain consistent across a wide range of demographic and structural assumptions, underscoring the methodological stability and general applicability of the proposed approach.

Practical recommendations

Based on the experimental evidence and practical implications of each strategy, the following guidelines are proposed:

- The heuristic assignment method is recommended in time-constrained or resource-limited environments, where simplicity, feasibility, and transparency are essential.
- Simulated annealing is better suited for scenarios that require high-quality and diverse solutions. Depending on the desired density and diversity of the Pareto front, SA can be initialized using either heuristic-based solutions or deliberately fragmented configurations.

FUTURE WORK

Several promising directions can be pursued to extend this research:

- **Integration of dynamic contagion models** (*e.g.*, Susceptible–Infectious–Recovered (SIR), Susceptible–Exposed–Infectious–Recovered (SEIR)) to simulate how structural configurations influence infection dynamics over time, moving beyond the current static exposure framework (*Holme & Saramäki, 2012*; *Pastor-Satorras et al., 2015*; *Aleta & Moreno, 2020*; *Borges et al., 2022*).
- **Validation using real-world data**, including actual school rosters, sibling registries, and contact-tracing records, to assess the generalizability and robustness of the proposed methods across diverse socioeconomic and cultural contexts beyond Spain.
- **Explicit modeling of non-reversible classroom-size heterogeneity.** While heterogeneity in enrollment does not pose a methodological limitation when classrooms can be rebalanced *a posteriori*—as is common when defining parallel classroom lines within a grade—some educational settings impose administrative, legal, or pedagogical constraints that prevent such rebalancing. In these cases, class-size heterogeneity becomes an endogenous and fixed feature of the system. Extending the framework to simulate enrollment-driven processes under non-adjustable capacity constraints will allow the evaluation of the robustness of the proposed strategies under more restrictive real-world conditions.
- **Relaxation of strict group-size constraints**, allowing for minor size imbalances between classrooms, could enhance the flexibility and performance of optimization strategies.
- **Incorporation of logistical and pedagogical constraints**, such as teacher availability, subject-specific scheduling, or physical classroom limitations, to improve the realism and applicability of the model.
- **Exploration of hybrid or trajectory-based optimization techniques**, including dynamic heuristics or memetic algorithms, which may offer better scalability and adaptability throughout the academic year.

This work provides a foundation for developing data-driven, equitable, and safe strategies for classroom structuring—particularly relevant in post-pandemic educational planning—and contributes both a formal framework and methodological tools for addressing complex scheduling problems under social and epidemiological constraints.

ACKNOWLEDGEMENTS

Special thanks go to Luis R. Izquierdo for his invaluable discussions that greatly contributed to this work, and to María Ojeda Ruiz and Rubén Arasti Blanco, former Computer Science students at the University of Burgos, for their implementation of this approach through various web applications. We acknowledge the Santander Supercomputación support group at the University of Cantabria for providing access to the Altamira Supercomputer at the Institute of Physics of Cantabria (IFCA-CSIC), a member of the Spanish Supercomputing Network, which was used to perform the simulations and analyses.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research was supported by the Ministry of Science and Innovation through its excellence network RED2022-134890-T, the project PID2020118906GB-I00, and the MOMENTUM program project MMT24-IMF-02. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Ministry of Science and Innovation through its Excellence Network: RED2022-134890-T, PID2020118906GB-I00.
MOMENTUM Program Project: MMT24-IMF-02.

Competing Interests

José Manuel Galán and José Ignacio Santos are both Academic Editors for PeerJ.

Author Contributions

- José Manuel Galán conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Silvia Díaz-de la Fuente conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Virginia Ahedo conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- José Ignacio Santos conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The datasets and code are available at GitHub and Zenodo:

- https://github.com/josemagalan/Siblings-Rewiring

- José Manuel, G., Silvia, D.-. de . la F., Virginia, A., & José Ignacio, S. (2026). Sibling Rewiring Project (1.1.0). Zenodo. https://doi.org/10.5281/zenodo.18411467

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.3710#supplemental-information.

REFERENCES

Aarts E, Korst J, Michiels W. 2005. Simulated annealing. In: Burke EK, Kendall G, eds. *Search Methodologies*. Boston, MA: Springer, US, 187–210.

Aleta A, Moreno Y. 2020. Evaluation of the potential incidence of COVID-19 and effectiveness of containment measures in Spain: a data-driven approach. *BMC Medicine* **18(1)**:157 DOI 10.1186/s12916-020-01619-5.

Alonso S, Alvarez-Lacalle E, Català M, López D, Jordan I, García-García JJ, Soriano-Arandes A, Lazcano U, Sallés P, Masats M, Urrutia J, Gatell A, Capdevila R, Soler-Palacin P, Bassat Q, Prats C. 2021. Age-dependency of the propagation rate of coronavirus disease 2019 inside school Bubble Groups in Catalonia, Spain. *Pediatric Infectious Disease Journal* **40(11)**:955–961 DOI 10.1097/INF.0000000000003279.

Al-Saleh MF, AL-Batainah FK. 2021. Estimation of the proportion of sterile couples using the negative binomial distribution. *Journal of Data Science* **1(3)**:261–274 DOI 10.6339/JDS.2003.01(3).130.

Benner AD. 2011. The transition to high school: current knowledge, future directions. *Educational Psychology Review* **23(3)**:299–328 DOI 10.1007/s10648-011-9152-0.

Block P, Hoffman M, Raabe IJ, Dowd JB, Rahal C, Kashyap R, Mills MC. 2020. Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nature Human Behaviour* **4(6)**:588–596 DOI 10.1038/s41562-020-0898-6.

Borges ME, Ferreira LS, Poloni S, Bagattini AM, Franco C, da Rosa MQM, Simon LM, Camey SA, de Kuchenbecker RS, Prado PI, Diniz-Filho JAF, Kraenkel RA, Coutinho RM, Toscano CM. 2022. Modelling the impact of school reopening and contact tracing strategies on COVID-19 dynamics in different epidemiologic settings in Brazil. *Global Epidemiology* **4**:100094 DOI 10.1016/j.gloepi.2022.100094.

Burgio G, Gómez S, Arenas A. 2023. Spreading dynamics in networks under context-dependent behavior. *Physical Review E* **107(6)**:064304 DOI 10.1103/PhysRevE.107.064304.

Casella G, Berger RL. 2002. *Statistical inference*. Pacific Grove, CA: Thomson Learning.

Casserly AM, Tiernan B, Maguire G. 2019. Primary teachers' perceptions of multi-grade classroom grouping practices to support inclusive education. *European Journal of Special Needs Education* **34(5)**:617–631 DOI 10.1080/08856257.2019.1580835.

Cave P. 2007. *Primary school in Japan*. Oxfordshire: Routledge.

Chami GF, Ahnert SE, Kabatereine NB, Tukahebwa EM. 2017. Social network fragmentation and community health. *Proceedings of the National Academy of Sciences of the United States of America* **114(36)**:E7425–E7431 DOI 10.1073/pnas.1700166114.

Coello Coello CA. 2006. Evolutionary multi-objective optimization: a historical view of the field. *IEEE Computational Intelligence Magazine* **1(1)**:28–36 DOI 10.1109/MCI.2006.1597059.

Cohen R, Havlin S, ben-Avraham D. 2003. Efficient immunization strategies for computer networks and populations. *Physical Review Letters* **91(24)**:247901 DOI 10.1103/PhysRevLett.91.247901.

Danon L, Lacasa L, Brooks-Pollock E. 2021. Household bubbles and COVID-19 transmission: insights from percolation theory. *Philosophical Transactions of the Royal Society B: Biological Sciences* **376(1829)**:629 DOI 10.1098/rstb.2020.0284.

Deb K, Pratap A, Agarwal S, Meyarivan T. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* **6(2)**:182–197 DOI 10.1109/4235.996017.

Esposito S, Cotugno N, Principi N. 2021. Comprehensive and safe school strategy during COVID-19 pandemic. *Italian Journal of Pediatrics* **47(1)**:6 DOI 10.1186/s13052-021-00960-6.

European Commission/EACEA/Eurydice. 2023a. *Organisation de l'enseignement primaire*. France: Eurydice–Eurypedia.

European Commission/EACEA/Eurydice. 2023b. *Organisation of primary education*. Germany: Eurydice–Eurypedia.

European Commission/EACEA/Eurydice. 2025. *Organización de la educación primaria*. España: Eurydice–Eurypedia.

Fehr E, Schmidt KM. 1999. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics* **114(3)**:817–868 DOI 10.1162/003355399556151.

Fukumoto K, McClean CT, Nakagawa K. 2021. No causal effect of school closures in Japan on the spread of COVID-19 in spring 2020. *Nature Medicine* **27(12)**:2111–2119 DOI 10.1038/s41591-021-01571-8.

Gallos LK, Liljeros F, Argyrakis P, Bunde A, Havlin S. 2007. Improving immunization strategies. *Physical Review E* **75(4)**:45104 DOI 10.1103/PhysRevE.75.045104.

Galán JM, Díaz-de la Fuente S, Ahedo V, Santos JI. 2026. Family-aware grouping strategy for infection control in schools. *Lecture Notes on Data Engineering and Communications Technologies* **276**:173–178 DOI 10.1007/978-3-032-10126-6_28.

Hallam S, Ireson J, Davies J. 2004. Grouping practices in the primary school: what influences change? *British Educational Research Journal* **30(1)**:117–140 DOI 10.1080/01411920310001629992.

Henderson D, Jacobson SH, Johnson AW. 2003. The theory and practice of simulated annealing. In: Glover F, Kochenberger GA, eds. *Handbook of Metaheuristics*. Boston: Kluwer Academic Publishers, 287–319.

Herbst M, Sobotka A, Wójcik P. 2023. The effect of peer group stability on achievements: evidence from Poland. *European Journal of Education* **58(1)**:166–180 DOI 10.1111/ejed.12545.

Herrmann HA, Schwartz JM. 2020. Why COVID-19 models should incorporate the network of social interactions. *Physical Biology* **17(6)**:065008 DOI 10.1088/1478-3975/aba8ec.

Holme P. 2004. Efficient local strategies for vaccination and network attack. *Europhysics Letters (EPL)* **68(6)**:908–914 DOI 10.1209/epl/i2004-10286-2.

Holme P, Saramäki J. 2012. Temporal networks. *Physics Reports* **519(3)**:97–125 DOI 10.1016/j.physrep.2012.03.001.

INE. 2020. *Encuesta continua de Hogares*. Instituto Nacional de Estadística. *Available at https://www.ine.es/*.

Ito H, Nakamuro M, Yamaguchi S. 2020. Effects of class-size reduction on cognitive and non-cognitive skills. *Japan and the World Economy* **53(3)**:100977 DOI 10.1016/j.japwor.2019.100977.

Jarosz B. 2021. Poisson distribution: a model for estimating households by household size. *Population Research and Policy Review* **40(2)**:149–162 DOI 10.1007/s11113-020-09575-x.

Jennings VE, Lloyd-Smith CW. 2015. Poisson and household size. *Mathematical Scientist* **40(1–2)**:103–117 DOI 10.1007/bf03029455.

Jennings V, Lloyd-Smith B, Ironmonger D. 1999. Household size and the poisson distribution. *Journal of the Australian Population Association* **16(1–2)**:65–84 DOI 10.1007/BF03029455.

Johnson NL, Kemp AW, Kotz S. 2005. *Univariate discrete distributions*. Hoboken: Wiley.

Kaiser SV, Watson A, Dogan B, Karmur A, Warren K, Wang P, Camano Sosa M, Olarte A, Dorsey S, Su M, Brown L, Sachdev D, Bardach NS. 2021. Preventing COVID-19 transmission in education settings. *Pediatrics* **148(3)**:1675 DOI 10.1542/peds.2021-051438.

Kamarul Aryffin HA, Bin Sahbudin MA, Ali Pitchay S, Abhalim AH, Sahbudin I. 2025. Technological trends in epidemic intelligence for infectious disease surveillance: a systematic literature review. *PeerJ Computer Science* **11(15)**:e2874 DOI 10.7717/peerj-cs.2874.

Kirkpatrick S, Gelatt CD, Vecchi MP. 1983. Optimization by simulated annealing. *Science* **220(4598)**:671–680 DOI 10.1126/science.220.4598.671.

Kiss IZ, Miller JC, Simon PL. 2017. *Mathematics of epidemics on networks*. Cham: Springer International Publishing.

Krishnaratne S, Pfadenhauer LM, Coenen M, Geffert K, Jung-Sievers C, Klinger C, Kratzer S, Littlecott H, Movsisyan A, Rabe JE, Rehfuess E, Sell K, Strahwald B, Stratil JM, Voss S, Wabnitz K, Burns J. 2020. Measures implemented in the school setting to contain the COVID-19 pandemic: a rapid scoping review. *Cochrane Database of Systematic Reviews* **2020(12)**:E1146 DOI 10.1002/14651858.CD013812.

Lai J, Luo T, Liu X, Huang L, Yu Z, Wang Y. 2023. Inferring household size distribution and its association with the built environment using massive mobile phone data. *Cities* **136(3)**:104253 DOI 10.1016/j.cities.2023.104253.

Lessler J, Grabowski MK, Grantz KH, Badillo-Goicoechea E, Metcalf CJE, Lupton-Smith C, Azman AS, Stuart EA. 2021. Household COVID-19 risk and in-person schooling. *Science* **372(6546)**:1092–1097 DOI 10.1126/science.abh2939.

Littlecott H, Krishnaratne S, Burns J, Rehfuess E, Sell K, Klinger C, Strahwald B, Movsisyan A, Metzendorf M-I, Schoenweger P, Voss S, Coenen M, Müller-Eberstein R, Pfadenhauer LM. 2024. Measures implemented in the school setting to contain the COVID-19 pandemic. *Cochrane Database of Systematic Reviews* **2024(5)**:E1027 DOI 10.1002/14651858.CD015029.pub2.

Liu Y, Huang Z, Wang G. 2023. Student learning performance prediction based on online behavior: an empirical study during the COVID-19 pandemic. *PeerJ Computer Science* **9(4)**:e1699 DOI 10.7717/peerj-cs.1699.

Lo Moro G, Sinigaglia T, Bert F, Savatteri A, Gualano MR, Siliquini R. 2020. Reopening schools during the COVID-19 pandemic: overview and rapid systematic review of guidelines and recommendations on preventive measures and the management of cases. *International Journal of Environmental Research and Public Health* **17(23)**:1–21 DOI 10.3390/ijerph17238839.

Maheshwari P, Albert R. 2020. Network model and analysis of the spread of COVID-19 with social distancing. *Applied Network Science* **5(1)**:509 DOI 10.1007/s41109-020-00344-5.

Marmot M. 2005. Social determinants of health inequalities. *The Lancet* **365(9464)**:1099–1104 DOI 10.1016/S0140-6736(05)71146-6.

McLeod RS, Hopfe CJ, Bodenschatz E, Moriske H, Pöschl U, Salthammer T, Curtius J, Helleis F, Niessner J, Herr C, Klimach T, Seipp M, Steffens T, Witt C, Willich SN. 2022. A multi-layered strategy for COVID-19 infection prophylaxis in schools: a review of the evidence for masks, distancing, and ventilation. *Indoor Air* **32(10)**:1 DOI 10.1111/ina.13142.

National Institute for Educational Policy Research (NIER). 2011. *Primary schools in Japan*. Tokyo: NIER.

Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. 2015. Epidemic processes in complex networks. *Reviews of Modern Physics* **87(3)**:925–979 DOI 10.1103/RevModPhys.87.925.

Patil GP, Rao CR. 1978. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics* **34(2)**:179 DOI 10.2307/2530008.

Rao BR, Mazumdar S, Waller JH, Li CC. 1973. Correlation between the numbers of two types of children in a family. *Biometrics* **29(2)**:271 DOI 10.2307/2529391.

Rockoff JE. 2004. The impact of individual teachers on student achievement: evidence from panel data. *American Economic Review* **94(2)**:247–252 DOI 10.1257/0002828041302244.

Sombetzki M, Lücker P, Ehmke M, Bock S, Littmann M, Reisinger EC, Hoffmann W, Kästner A. 2021. Impact of changes in infection control measures on the dynamics of COVID-19 infections in schools and pre-schools. *Frontiers in Public Health* **9**:780039 DOI 10.3389/fpubh.2021.780039.

Soriano-Paños D, Cota W, Ferreira SC, Ghoshal G, Arenas A, Gómez-Gardeñes J. 2022. Modeling communicable diseases, human mobility, and epidemics: a review. *Annalen der Physik* **534(6)**:11001 DOI 10.1002/andp.202100482.

Sundaram N, Bonell C, Ladhani S, Langan SM, Baawuah F, Okike I, Ahmad S, Beckmann J, Garstang J, Brent BE, Brent AJ, Amin-Chowdhury Z, Aiano F, Hargreaves J. 2021. Implementation of preventive measures to prevent COVID-19: a national study of English primary schools in summer 2020. *Health Education Research* **36(3)**:272–285 DOI 10.1093/her/cyab016.

Talbi E-G. 2009. *Metaheuristics: from design to implementation*. Hoboken, NJ: Wiley.

Tricomi E, Rangel A, Camerer CF, Odoherty JP. 2010. Neural evidence for inequality-averse social preferences. *Nature* **463(7284)**:1089–1091 DOI 10.1038/nature08785.

Vespignani A, Tian H, Dye C, Lloyd-Smith JO, Eggo RM, Shrestha M, Scarpino SV, Gutierrez B, Kraemer MUG, Wu J, Leung K, Leung GM. 2020. Modelling COVID-19. *Nature Reviews Physics* **2(6)**:279–281 DOI 10.1038/s42254-020-0178-4.

Wellington A, Clark M, Burnett A, James-Burdumy S, Makowsky L, Brockman S, Dotter D, Herrmann M, Chiang H. 2024. *Evaluation of departmentalized instruction in elementary schools: exploring implementation experiences (NCEE 2024-005)*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Wentzel KR. 1999. Social-motivational processes and interpersonal relationships: implications for understanding motivation at school. *Journal of Educational Psychology* **91(1)**:76–97 DOI 10.1037/0022-0663.91.1.76.

Zitzler E, Deb K, Thiele L. 2000. Comparison of multiobjective evolutionary algorithms: empirical results. *Evolutionary Computation* **8(2)**:173–195 DOI 10.1162/106365600568202.