An extended systematic mapping study about
the scalability of i* Models

Paulo Lima¹, Jéssyka Vilela¹, Enyo Gonçalves¹², João Pimentel³, Ana Holanda¹⁴, Jaelson Castro¹, Fernanda Alencar¹, Maria Lencastre⁵
¹Universidade Federal de Pernambuco (UFPE)
²Universidade Federal do Ceará (UFC)
³Universidade Federal Rural de Pernambuco (UFRPE)
⁴Instituto Federal do Acre (IFAC)
⁵Universidade de Pernambuco (UPE)
{plc2, jffv, ejtg, jhcp, acah, jbc}@cin.ufpe.br
fernandaalenc@gmail.com
mlpm@ecomp.poli.br

Abstract

i* models have been used for requirements specification in many domains, such as healthcare, telecommunication, and air traffic control. Managing the scalability and the complexity of such models is an important challenge in Requirements Engineering (RE). Scalability is also one of the most intractable issues in the design of visual notations in general: a well-known problem with visual representations is that they do not scale well. This issue has led us to investigate scalability in i* models and its variants by means of a systematic mapping study. This paper is an extended version of a previous paper on the scalability of i* including papers indicated by specialists. Moreover, we also discuss the challenges and open issues regarding scalability of i* models and its variants. A total of 126 papers were analyzed in order to understand: how the RE community perceives scalability; and which proposals have considered this topic. We found that scalability issues are indeed perceived as relevant and that further work is still required, even though many potential solutions have already been proposed. This study can be a starting point for researchers aiming to further advance the treatment of scalability in i* models.

Keywords: Systematic Mapping study, Scalability, Models, iStar, i*, Goal Models.

1 Introduction

The i* framework is a modeling language used in Requirements Engineering (RE) to create models that represent stakeholders, systems, and their dependencies [1]. Requirements are represented by elements of interest, namely goals, softgoals, resources and tasks. The i* language has been used in several situations [1], such as telecommunications, air traffic control, agriculture, e-government, healthcare and business process.

Unfortunately, i* is not suitable for modeling complex cases or involving many parts [20]. The limitation in scaling i* models is identified as one of their biggest barriers for industrial adoption [1][20]. If there were improved support for creating good i* models of large and complex systems, its adoption would be facilitated. Therefore, the i* framework requires solutions and means to address its scalability [3][20].

In a previous paper [1], we conducted a systematic mapping study to identify the primary studies on the scalability of i* models and its variants, following a predefined review protocol. In this paper, we present an extended version of the study including papers indicated by specialists, while also discussing the challenges and open issues regarding scalability of i* models and its variants. We identified 126 papers about i* scalability and analyzed the distribution of these studies, definitions, artefacts that aim to address it; how the researchers’ community perceive the i* ability to scale, and which are the open issues related to this theme.
The results of this paper are important to support researchers in the \( i^* \)/RE community since \( i^* \) models can become complex and could be difficult to decompose into smaller pieces [20]. The list of papers that mentioned \( i^* \) models scalability (or any of its related attributes) may serve to aid future systematic reviews. The challenges described in this paper are useful to identify emerging trends and provide an overall view of the problems tackled in the literature.

This paper is structured as follows: Section 2 discusses related work; the study design is presented in Section 3; the results related to each research question are reported in Section 4; Section 5 presents the discussion of the results; and Section 6 shows the conclusions and future work of this paper.

2 Related Work

The scalability of \( i^* \) is an issue addressed by many researchers. However, we identified only two secondary studies that synthesize \( i^* \) and its scalability: [3] outlined published research on \( i^* \) in general, commenting also the question of scalability, and [20] present many great challenges and \( i^* \) problems, including scalability notation.

The first work [3] cites some approaches proposed to improve the scalability. This work reviews the approaches adopted by the \( i^* \) framework and reports their application in various fields. The authors of this paper explained that models need to reflect software in their social characteristics of complex systems. They argue that \( i^* \) has a scalability limitation. The second work [20] exposed the scalability as one of the great challenges of \( i^* \). It argues that scalability and complexity are two characteristics that \( i^* \) does not support well and proposed a set of concepts to help to improve it.

In order to understand the progress that has been made in integrating goal models with downstream system development, [16] conducted a systematic survey to investigate what approaches exist which map/integrate/transform goal-oriented languages to other software artefacts or languages. They produced a roadmap summarizing 174 publications. They also comment the lack of widespread industrial adoption of goal-oriented models that could be attributed to several factors including the scalability and usability challenges in complex models, which are not easily decomposable.

Horkoff and Yu [4] use three available tools implementing seven similar analysis procedures for goal satisfaction to analyze three sample goal models. They performed a comparison to understand the ways in which procedural design choices affect analysis results, and how differences in analysis results could lead to different recommendations over alternatives in the model.

After analyzing these studies, we can notice that scalability is a desirable quality. However, this question is addressed in a broad sense and it becomes difficult to categorize possible evidence since it is not possible to replicate those results. Therefore, it is necessary to know all the works published in a more systematic way. Accordingly, our work presents the results of an extended systematic mapping study conducted to investigate the scalability of \( i^* \). We analyzed the distribution of these studies, definitions, mentions for artefacts to address it; how the researchers community perceive the \( i^* \) ability to scale and, which are the open issues related to this theme. In the next section, we describe our research protocol.

3 STUDY DESIGN

The purpose of this paper is to map the studies that addressed scalability issues in \( i^* \) and its variants. Hence, the search string was designed to only obtain papers about \( i^* \) or its variants (GRL, Tropos), excluding any other goal modelling languages (such as KAOS, NFR, etc). This type of research provides a summary of evidence related to a specific intervention strategy, by applying explicit methods and systematic search, critical appraisal and synthesis of selected information [6].

We conducted a systematic mapping study following the guidelines of Kitchenham and Charters [5]. We also consulted recently published systematic mappings in well-known Software Engineering Journals such as Journal of Systems and Software, Empirical Software Engineering Journal, and Requirements Engineering Journal.

The study consists of the following steps: (1) identification of the need for a systematic mapping; (2) formulation of focused research questions; (3) comprehensive, exhaustive search for primary studies; (4) identification of the data needed to answer the research questions; (5) data extraction; (6) summary and synthesis of study results; (7) interpretation of the results to determine their applicability; and (8) report-writing.

The study design gathers best practices of the empirical software engineering community. Experienced researchers validated it and it was adjusted through their feedback.
3.1 RESEARCH QUESTIONS

This paper aims to answer the following main research question:

*Which are the published studies that mention scalability of i* models in the requirements engineering context?*

Based on the main research question, specific questions were raised according to aspects of interest.

- RQ1: What studies mention the issue of i* models scalability?
- RQ2: What are the scalability definitions in the context of i* models?
- RQ3: What types of contributions have been published to support the scalability of i* models?
- RQ4: What judgments exist about the scalability of i* models?
- RQ5: What open issues have been identified regarding the scalability of i* models?

The first research question (RQ1) focuses on works that address the scalability or similar attributes such as modularity and complexity of i* models and its variants. The second research question (RQ2) relates to conceptualizations of scalability. The third research question (RQ3) quantifies the contributions, such as metamodels, formalisms, modelling processes, visual constructors, software and algorithms or other artefacts. The fourth research question (RQ4) summarizes published findings on the scalability of i* models being well treated or not. The fifth research question (RQ5) summarizes evidences of the open issues mentioned in the selected studies. Therefore, this paper aims to synthesize information on i* models regarding their scalability.

3.2 EXCLUSION AND INCLUSION CRITERIA

The selected studies were primary publications that mentioned the property of i* model scalability, present any discussion about or study the scalability of i* models, i.e. the research object was i* models. The exclusion and inclusion criteria adopted are presented in Table 1 and Table 2 respectively.

<table>
<thead>
<tr>
<th>ID</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC01</td>
<td>Studies not captured by the keywords in search engines.</td>
</tr>
<tr>
<td>EC02</td>
<td>Studies published before 1990.</td>
</tr>
<tr>
<td>EC03</td>
<td>Unreadable files (files corrupted).</td>
</tr>
<tr>
<td>EC04</td>
<td>Studies not written in English.</td>
</tr>
<tr>
<td>EC05</td>
<td>Studies that do not mention i* or variants.</td>
</tr>
<tr>
<td>EC06</td>
<td>Studies that do not mention scalability (and similar terms) of i* or variants</td>
</tr>
<tr>
<td>EC07</td>
<td>Non-Scientific studies (notes, index, editorials, prefaces).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC01</td>
<td>Studies that were not eliminated by the exclusion criteria.</td>
</tr>
<tr>
<td>IC02</td>
<td>Texts published between 01/01/1990 and 31/12/2014.</td>
</tr>
<tr>
<td>IC03</td>
<td>Studies that address some of the study questions</td>
</tr>
<tr>
<td>IC04</td>
<td>Theoretical or empirical work will be included.</td>
</tr>
</tbody>
</table>

3.3 SOURCES SELECTION AND SEARCH

The search strategy included only electronic databases and was validated by experts on the RE area. By using a search string, the following electronic databases were automatically searched: Science Direct, ACM Digital Library, IEEE Xplore, Engineering Village, Scielo and World Scientific. Figure 1 shows the systematic review process.

We developed the search string by specifying the main terms used about this topic, the derivation form constructors and synonyms arising from the research questions, previously readings from known studies, consulting from experienced researchers in the field and dictionaries or glossaries.
We performed pilot searches to refine the keywords in the search string using trial and error. After some iterations, we settled on the following search string:

("iStar" OR "i-star" OR "i star" OR "Yu, e" OR "Yu e" OR "GRL" OR "Tropos") AND ("goal-oriented" OR "goal-directed" OR "agent oriented" OR "requirements engineering" OR "software requirements") AND (model* OR diagram*) AND (scal* OR modul* OR complex* OR compreh* OR underst* OR evolu* OR large OR huge OR big OR siz* OR enormous OR immense); Publication Year: 1990 – 2014

Besides i* itself, its variants were also included in the search query: GRL and Tropos. Since scalability is a very specific term, the search query also contains attributes that are known to be related to scalability: modularity, complexity, comprehension, understandability, evolution, and size.

Considering that each search engine has their own syntax to perform automatic searches, the above search string had to be adapted to each one of them. After applying these steps, a total of 119 studies met the inclusion criteria and their data were extracted.

After consulting experienced researchers, they pointed out additional seven different studies that were not initially captured by our systematic mapping. These researchers have more than 10 years of experience in requirements engineering, and two of them have more than 20 years of experience in the field.

Considering the experience and knowledge of the researchers that recommended these studies, as well as the relevance of the content of these documents for this survey, they were included in this extended paper: [20], [22], [23], [24], [25], [26], [27], [28]. We believe that these studies were not captured by the search engines because some of them are PhD thesis and as such would not be the indexed in different search engines used, or not be indexed at all. The focus of our systematic mapping study was not to address these, therefore, we did not search in databases related to these kind of manuscripts.

Given that we manually added these papers, it was not necessary to change the search string. Hence, our mapping study relies on 126 papers regarding the scalability of i* models. The extraction was performed aiming to answer the research questions described in Section 3.1 The list of selected papers is presented in the appendix. In the next subsection we discuss some threats to the validity of this study and how they were handled.

### 3.4 Threats to Validity

This section describes some threats that must be improved in future replications of this study and other aspects that must be taken into account in order to generalize the results described in this paper [8]. We will base our discussion of threats to validity on the categories used by Wohlin et al. [7], but we only detail the categories that are considered important for our study.

Internal validity was enhanced using triangulation in some parts of the method. We consulted experienced researchers to validate the research design and its understanding. Their feedback and the trials contributed to reduce threats.
The construct validity required some extra care. This was necessary since the term scalability is an abstract word that has many definitions and the word \( i^* \) is difficult to insert in the search engines. Hence, to minimize threats of this nature, we discussed synonyms, written alternatives for both terms (for example, iStar and i-star for \( i^* \)). We also related potential previous definitions of scalability under the supervision of experienced researchers.

External validity is concerned with establishing the generalizability of the systematic mapping results, which is related to the degree to which the primary studies are representative of the reviewed topic [8]. The external validity (portability, transference) of this study was strengthened by the structure of the extracted data. It was also supported by detailing the research method in order to allow future comparative generalizations.

Regarding the empirical reliability, we tried to run a systematic mapping following protocols and studies already published and accepted by the academic community.

The focus of our systematic mapping study was not to address PhD thesis, therefore, we did not search databases related to these kind of manuscripts. Nonetheless, we believe that they can provide valuable results and it is necessary to consider them more systematically in future works.

Finally, systematic mapping studies and systematic literature reviews suffer with the issue of the coverage of selected studies returned with a search string. Hence, it is possible that the search string could omit relevant papers. To mitigate this issue, we consulted experienced researchers. From the studies pointed out by them, only 2 documents [26][27] could be returned by databases searched.

4 RESULTS

The purpose of this research is to know, in the context of Requirements Engineering, the publications about the \( i^* \) framework related to the scalability of this language and its variants. In the next sections, we present the answers for our research questions.

This paper is a systematic mapping study; as such, it focuses on quantitative data. Accordingly, it is not practicable to provide tables or to list all papers that answered each research question. This is a first work in the direction of understanding and solving such issue in the future.

RQ1: What studies mention the issue of \( i^* \) models scalability?

We found 126 studies that mention \( i^* \) models scalability or related attributes such as modularity, complexity, among others. Considering the research types proposed by [9] and also used by [8], we considered that 45 studies are more empirical, i.e. they were classified as evaluation research or validation research (see Figure 2).

![Figure 2: Selected studies distributed over quantity, year of publication and research categories.](image)
According to [9], evaluation research is the investigation of a problem in RE practice or an implementation of an RE technique in practice. This research type correspond to empirical studies, such as by case study, field study, field experiment, survey, etc.

Validation research, on the other hand, correspond to the investigation, using a thorough, methodologically sound research setup, of the properties of a solution proposal that has not yet been implemented in RE practice [9]. Possible research methods are experiments, simulation, prototyping, mathematical analysis, mathematical proof of properties, etc.

Solution proposal is a category of papers that describes a novel technique or at least a significant improvement of an existing technique [9]. Philosophical papers sketch a new way of looking at things, a new conceptual framework, etc. Opinion papers contain the author’s opinion about what is wrong or good about something, how we should do something, etc. Finally, experience papers contain a list of lessons learned by the author from his or her experience [9].

We can notice that the number of studies has increased since 2005, with peak of 19 publications in 2011. Figure 3 shows the relationship between the search sources and the number of excluded and included studies.

![Figure 3: Number of included and excluded studies per search sources.](image)

**RQ2: What are the scalability definitions in the context of i* models?**

From the selected studies, we identified 10 works that have some definition of the term scalability. These definitions are shown in Table 3. In addition, 11 studies had, in their research objective, the study of scalability or related attributes, such as modularity for example.

<table>
<thead>
<tr>
<th>Study</th>
<th>Definition or Characterization</th>
</tr>
</thead>
<tbody>
<tr>
<td>[53]</td>
<td>“Scalability was defined by the number of goal levels and number of variants.”</td>
</tr>
<tr>
<td>[89]</td>
<td>“able to have models at different levels of abstraction so that both domain experts and developers alike can get an idea of the overall system behavior or focus on a particular part of the system in more detail if required.”</td>
</tr>
<tr>
<td>[93]</td>
<td>“is able to handle numerous Agents in an application.”</td>
</tr>
<tr>
<td>[105]</td>
<td>“the property of reducing or increasing the scope of methods, processes, and management according to the problem size ( . . . ) Inherent in this idea is that software engineering techniques should provide good mechanisms for partitioning, composition, and visibility control. It includes the ability to scale the notation to particular problem needs, contractual requirements, or even to budgetary and business goals and objectives.”</td>
</tr>
<tr>
<td>[111]</td>
<td>“the reduced complexity of goal graphs ( . . . ), the ability to group goal graphs with concerns, the encapsulation provided by concerns, the ability to use parameterized point cut expressions in AoGRL, and</td>
</tr>
<tr>
<td>Study</td>
<td>Definition or Characterization</td>
</tr>
<tr>
<td>-------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>[124]</td>
<td>“measures the methodology’s support for designing systems that are scalable. It means that the system should allow the incorporation of additional resources and software components with minimal user disruption.”</td>
</tr>
<tr>
<td>[125]</td>
<td>“The degree to which the modelling framework can be used to handle applications of different sizes. Scalability also measures extensibility, the degree to which the inclusion of new modelling elements leaves the understandability of models unaffected. This feature is causally related to refinement and modularity.”</td>
</tr>
<tr>
<td>[129]</td>
<td>“features in the technique to scale with the size and complexity of the system under assessment. Examples: Abstraction, refinement, decomposition, different formats, types or versions of technique.”</td>
</tr>
<tr>
<td>[131]</td>
<td>“ability of both the approach as well as the specifications to serve for a variety of project sizes and constraints, need to be easily modifiable.”</td>
</tr>
<tr>
<td>[138]</td>
<td>“large organizational models (depending on the domain and their description) become complex and inconsistent due to bad labeling and irrelevant information.”</td>
</tr>
</tbody>
</table>

RQ3: What types of contribution have been published to support the scalability of i* models?

From the set of selected studies, we identified 150 references to different types of contributions to improve the scalability of i* models (Figure 4). This includes repetitions to the same technique as well as more than one type of contribution in the same study.

We separate contributions in four categories: metamodels or formalisms (24 mentions), modelling processes (54 mentions), visual constructors (44 mentions), and software or algorithms (34 mentions).

![Figure 4: Types of contributions per type of research method.](image)

RQ4: What judgments exist about the scalability of i* models?

According to the selected studies, we identified that i* models do not have a good treatment regarding their scalability. In all research types categories (Figure 5), there were quantitatively more studies that reported bad impression on the i* model supporting scalability (67 studies). On the other hand, only eight studies judged the scalability of i* model as being well treated. Finally, 50 studies did not mention any information about this question.
In relation to the research types categories (Figure 5), 5 studies that belong to the Validation category, 1 study from Evaluation and 2 studies from Solution Proposal category say that scalability is well treated. Besides, 8 studies from the Evaluation, 18 studies from Validation, 29 from Solution Proposal, 3 Opinion Papers, and 4 Experience category classified scalability of $i^*$ models as not well treated.

RQ5: What open issues have been identified regarding the scalability of $i^*$ models?

The aim of this question is to identify open issues reported in the selected studies regarding to scalability of $i^*$ models and its variants. We identified 100 studies that presents open issues and related future works about the scalability of $i^*$ models. Regarding the research types categories (Figure 6), 51 studies indicate open issues from Solution Proposal, 24 from Validation, 13 from Evaluation, 7 studies from Experience and 4 from Opinion papers category. On the other hand, 26 studies did not mention open issues.

The selected studies point out many open issues about the topic. The studies identified sixteen challenges, listed in Table 4 along with the number of papers that mention the challenge. The last column in Table 4 presents the percentage of papers that mentioned that challenge with respect to total number of selected papers. From the total of selected studies, 20.63% (26 studies) do not explicitly indicate any open issue about the investigated topic.
Table 4: Challenges and needs in scalability of $i^*$ models.

<table>
<thead>
<tr>
<th>Id</th>
<th>Challenges about scalability</th>
<th>#mentions</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>How to decrease the complexity of $i^*$ models.</td>
<td>40</td>
<td>31.75</td>
</tr>
<tr>
<td>C2</td>
<td>How to improve the scalability of $i^*$ models.</td>
<td>31</td>
<td>24.60</td>
</tr>
<tr>
<td>C3</td>
<td>How to increase the modularity of $i^*$ models.</td>
<td>18</td>
<td>14.29</td>
</tr>
<tr>
<td>C4</td>
<td>How to manage very large amount of information in $i^*$ models.</td>
<td>18</td>
<td>14.29</td>
</tr>
<tr>
<td>C5</td>
<td>How to define different views (perspectives) in $i^*$ models and maintain traceability.</td>
<td>14</td>
<td>11.11</td>
</tr>
<tr>
<td>C6</td>
<td>How to better understand the $i^<em>$ framework so that requirements engineers can fully explore $i^</em>$ strengths and build less complex $i^*$ models.</td>
<td>8</td>
<td>6.35</td>
</tr>
<tr>
<td>C7</td>
<td>Rudimentary available tool support.</td>
<td>8</td>
<td>6.35</td>
</tr>
<tr>
<td>C8</td>
<td>Definition of metrics to evaluate the complexity, scalability or correctness of $i^*$ models.</td>
<td>5</td>
<td>3.97</td>
</tr>
<tr>
<td>C9</td>
<td>How to improve the reuse and evolution of $i^*$ models.</td>
<td>5</td>
<td>3.97</td>
</tr>
<tr>
<td>C10</td>
<td>Usability or scalability evaluation of tools.</td>
<td>3</td>
<td>2.38</td>
</tr>
<tr>
<td>C11</td>
<td>Evaluate the scalability or modularity of the proposal.</td>
<td>3</td>
<td>2.38</td>
</tr>
<tr>
<td>C12</td>
<td>How to improve the visualization (legibility) of $i^*$ models.</td>
<td>3</td>
<td>2.38</td>
</tr>
<tr>
<td>C13</td>
<td>How to decrease the ambiguity of $i^*$ models.</td>
<td>2</td>
<td>1.59</td>
</tr>
<tr>
<td>C14</td>
<td>Development of a tool to support the proposal.</td>
<td>2</td>
<td>1.59</td>
</tr>
<tr>
<td>C15</td>
<td>How to reduce the learning curve of the $i^*$ modeling language.</td>
<td>2</td>
<td>1.59</td>
</tr>
<tr>
<td>C16</td>
<td>Evaluate the use of $i^*$ models in real projects.</td>
<td>2</td>
<td>1.59</td>
</tr>
<tr>
<td>SUM:</td>
<td></td>
<td>164</td>
<td></td>
</tr>
</tbody>
</table>

Although $i^*$ models have been used to specify different domains, the challenges listed in Table 4 are common to multiple domains, which indicates the difficulty to address the scalability issues of these models. Besides, these issues have only been partially addressed by the selected studies.

We noticed that the studies use the terms complexity and understandability as synonyms of scalability. The adequate treatment of complexity is the open issue most cited in the studies, being mentioned more than the scalability itself. Moreover, the concepts of modularity and visualization are also described as associated with the scalability concept.

$i^*$ models are designed for communication with end users, therefore it is especially important that the notations have effective complexity management mechanisms as non-experts are less equipped to deal with complexity [108]. Moreover, due to the high complexity of social relations, $i^*$ models may fail to cover all relevant issues [127]. The authors of [142] complement stating that as with other modeling techniques, in choosing to highlight certain features of a complex reality, many other aspects are omitted. The complexity of the social world presents a formidable challenge for a modeling approach. Hence, how to decrease the complexity of the $i^*$model (C1) is an open issue cited by 40 studies (31.75%).

It is well known that visual representations do not scale well. As a consequence, $i^*$ models soon become very large, often with many links between elements, which users find confusing [92]. Therefore, as expected, scalability is recognized to be missing aspect [111]. The results confirmed that how to improve the scalability of $i^*$model (C2) is a challenge well cited in the studies (31 papers; 24.60%).

An alternative to improve the scalability and reduce the complexity is pointed out by 18 studies (14.29%) through increasing the modularity of $i^*$ models. However, the studies also affirm that $i^*$ models lack modularity. This has a specially negative impact for development, since requirement models tend to include a huge number of elements with crossed relationships between them. In turn, the readability of the models is decreased, harming their utility and increasing the error rate and development time [105]. Hence, a challenge is how to increase the modularity of...
i*model (C3) by dividing the model into small pieces avoiding scalability problems and also improving the stakeholders’ understanding [108].

i* models do not support the management of very large amount of information [94]. Hence, how to manage very large amount of information in i*model (C4) is indicated by 18 studies (14.29%) as an open issue. The challenge C4 is related to the open issue regarding how to define different views (perspectives) of i*model and maintain traceability (C5) mentioned by 14 studies (11.11%). According to [144], because of the multiplicity and complexity of organizational issues, many authors have advocated the need for multiple perspectives when trying to understand organizational phenomena. Understanding an organizational setting often require a number of perspectives. Given the complexity of issues within one perspective, it is difficult to bring multiple perspectives together to draw conclusions from their combined insights.

The number of elements in i*models that can be comprehended at a time is limited by working memory capacity (believed to be seven, plus or minus two elements at a time). When this is exceeded, a state of cognitive overload ensues and comprehension degrades rapidly [26]. Moreover, considering that building any diagram is more difficult than reading it [120], how to better understand the i* framework so that requirements engineers can fully explore i* strengths and build less complex i* model is a challenge in 8 studies (6.35%).

It is reasonable to assume that a considerable effort is needed to produce a complete SR model for even a moderately complex system, especially in large-scale systems. As well as the intellectual effort needed to develop each actor model, the resulting SR models were large and somewhat difficult to manage, especially given the rudimentary tool support available (C7) [100]. This lack of adequate tools is mentioned by 8 studies (6.35%) as an open issue.

The Definition of metrics to evaluate the complexity, scalability or correctness of i*model (C8) is a concern of 5 studies (3.97%). The i* community should get involved in this complex activity [46] to propose these metrics [31].

Reuse and evolvability is an important feature of software. Software systems evolve continuously in order to fulfil new requirements of stakeholders, and to adapt changing business rules and environments [65]. The studies indicate that i*models are difficult to evolve or reuse [29], and therefore, How to improve the reuse and evolution of i*model (C9) is cited by 5 studies (3.97%) as a challenge for the adoption of i*models and for their scalability.

The Usability or scalability evaluation of tools (C10) is another important aspect highlighted in 3 studies (2.38%). Tools should be easy to use and scalable to reduce the cognitive complexity and the mental burn during the modeling.

Evaluate the scalability or modularity of the proposal (C11), mentioned in 3 studies (2.38%), is fundamental for the application of i*models in real projects models, wide spread adoption of the notations and use by the industry. A related open issue is Evaluate the use of i* model in real projects (C16) cited in 2 studies (1.59%).

Modelers are forced to expend considerable effort considering models in a legible way and attempting to trace dependencies from one to another. The effort and time spent on such tasks detracts from the central aim of the models [145]. Moreover, the size of the resulting i * model can be a challenge in terms of visualization, and ways of exploring dynamically a large i*model are hence needed [128]. This open issue of How to improve the visualization (legibility) of the models (C12) is mentioned in 3 studies (2.38%).

There is a need for formal gathering mechanisms, which provide the necessary degree of non-ambiguity and detail [106]. These mechanisms can contribute to solve the problem of How to decrease the ambiguity of i* models (C13) cited by 2 studies (1.59%).

Many proposals for improving the scalability of i* models do not have tool support. Hence the development of a tool to support the proposal (C14) is explicitly pointed out by 2 studies (1.59%) is an open issue.

Lastly, the learning curve of i* model is considered high by 2 studies (1.59%) making it difficult to build models that are easier to read and understand.

Considering the distribution of the challenges according to publication year, some challenges have not been addressed in recent years: C8, C9, and C13 (last mention in 2010); C16 (last mention in 2009); C14 (last mention in
Challenges C14 and C15 were only mentioned twice in the same year, showing that there were no advances regarding these challenges throughout the years. The first challenge to be mentioned was C6, in 1997, followed by C2 and C3 in 2000. The challenges that have been mentioned for the first time most recently were C11 and C15, both in 2011. Figure 7 depicts the number of papers that mention each of the top five challenges on each year. It can be noted that none of these challenges has been abandoned in recent years.

![Figure 7: Number of papers mentioning the top 5 challenges throughout the years.](image)

## 5 Discussion

In this Section, we first discuss all research questions and their answers. Then, we report on the scalability treatments of some of the reviewed papers. Last but not least, we present some general limitations and recommendations.

The first question is related to the selection of studies that mention the issue of $i^*$ model scalability or some associated quality attribute. The list of 126 studies that mention this issue is an expressive number that can be used for any research that aims to understand such attribute in a detailed way. The papers were classified according to their research type: philosophical, experience, opinion, solution proposal, validation, and evaluation. Therefore, a researcher can use our results to obtain papers related to her research agenda. In particular, it was observed that 35.7% of the studies found are empirical, showing a reasonable maturity of this research community.

The second question addressed the scalability definitions in the context of $i^*$ model. We provided a list of 10 definitions for $i^*$ models scalability extracted from the analyzed papers (Table 3). That list contributes to the understanding of this quality attribute. It also allows the proper characterization of this attribute by aligning it with existing general definitions of scalability or by creating a new one that encompasses those that are listed here. In particular, it was observed that some definitions are focused on the models themselves, whereas others focus on methodological support.

The third research question aimed at investigating the types of contributions that have been published to support the $i^*$ model scalability. We found more than 150 contributions in four categories: metamodels or formalisms (24 mentions), modeling processes (54 mentions), visual constructors (44 mentions), and software or algorithms (34 mentions). We observed that the distribution of these categories is very similar in different types of research, most notably: evaluation research, validation research, and solution proposal. By classifying the papers in different kinds of contributions, this work promotes the reuse of such solutions. Researchers may analyze the existing solutions to avoid rework. Finally, new solutions should be compared against the current ones.

There are many different judgments about the scalability of $i^*$ models (research question 4). These judgments allow to understand how the topic is perceived by its research community. A total of 67 studies report that $i^*$ does not appropriately support scalability. In contrast, 8 studies consider the scalability quality attribute as being already well
treated. This mapping study is a first attempt to analyze the judgment regarding this topic. We could conclude that the majority of studies disapprove how the scalability is addressed in the i* framework. Nonetheless, some studies (including studies published recently) consider it well treated. Therefore, this issue deserves a more thorough investigation through systematic literature reviews or through the conductions of experiments in order to analyze specific aspects of the issue. For example, is there a maximum size of elements for a model to be considered understandable? Are there different representations that can be adopted for large models?

Finally, the last research question investigated the open issues that have been identified regarding the scalability of i* models. We found a total of 100 studies that present open issues and related future works about the scalability of i* models. The challenges most cited were: how to decrease the complexity of i* models; how to improve the scalability of i* models; how to increase the modularity of i* models; how to manage very large amounts of information in i* models; and how to define different views (perspectives) in i* models while maintaining traceability. Moreover, it was cited the need for the development of modeling tools as well as the necessity to conduct more experiments to obtain more evidence. Finally, many studies report open issues regarding their own research, not directly related to scalability – therefore, they were not considered.

As a summary, the following limitations on the overall research about i* scalability were identified: small amount of evaluation research on the topic; scarcity of i* models from the industry; there are many contributions on the topic, but no clear recommendations on which approach to adopt on different contexts; lastly, some relevant publications are not readily available, preventing wider adoption of the proposed approaches.

The following subsection presents some of the mechanisms to handle complexity that were found in this systematic mapping study.

5.1 THREATS TO VALIDITY

There are many different ways to handle the scalability of i* models, some of which are briefly described in this subsection. With this, we expect to provide an overview of the kinds of mechanisms that have been proposed by this research community.

Maté et al [15] defined modules for i* models. On an empirical evaluation comparing between regular models and models with modules, the authors observed increased scalability and improved understandability (considering the number of errors) on the latter, even though it took longer to understand them.

Pastor et al [14] presented an empirical evaluation of i* with respect to different concerns, including modularity, complexity management, and scalability. Both modularity and scalability were not considered supported, whereas complexity management was not considered to be well supported.

Alencar et al [10] extended i* with aspectual elements aiming to improve the modularity of its models. By distributing repeating elements on aspectual actors, it is possible to increase the separation of concerns, with the disadvantage of being necessary to learn new elements in order to use this approach. With increased modularity and separation of concerns, the models are expected to be more scalable.

Previously, Mussbacher et al [111] have also investigated the adoption of aspectual concepts in the context of i* based models. The proposal was evaluated by comparing three approaches for modelling an example system: regular GRL; monotonic GRL; and aspect-oriented GRL. The latter presented better results in terms of modularity, understandability, reusability, and maintainability, even though it increased the vocabulary size.

Oliveira et al [13] defined SDsituations, a modularity construct explicitly aimed at improving the scalability of i* models. SDsituations aggregate different elements of i* models, and each SDsituation is related to other SDsituations through logical, temporal or sequential, and physical dependencies.

The approach from Dalpiaz et al [11] allows designing adaptive socio-technical systems. Here, instead of considering the scalability of the models themselves, it was analyzed the scalability of the adaptation algorithms that take i* models as input.

Similarly, Aydemir et al [12] assessed the scalability of their algorithms for model evolution. Moreover, it explicitly takes some precautions in order to improve its visual scalability, such as high visual distance between different kinds of elements and one-to-one correspondence between symbols and concepts.

Horkoff and Yu [75] address the problem of performing an interactive analysis of large goal models by highlighting specific elements depending on the current analysis step.

iStar was extended in [17] to make explicit crosscutting concerns, as a means to address more efficiently requirements change and its impact on other requirements. A new graphical representation in the form of a star has
been proposed to represent the crosscutting application requirements and avoid duplication in the representation of these models, improving scalability.

Links to other parts of the diagram are also used to improve the modularity and scalability of i* models. For example, a link can be used to indicate a part of the model. Hence, an i* diagram can be split in parts. In Ali et al. [18], links to other parts of the diagram are used to detail context in i* relationships.

Finally, URN (User Requirement Notation) [19] is a fusion of iStar with use case maps. This modeling language has entities to modularize its models and improve the scalability of its models.

6 Conclusions and Future Works

In this paper we have presented an extended systematic mapping on the scalability of i* models and its variants. From a initial set of 2774 papers, after applying the inclusion and exclusion criteria a total of 119 papers were selected for further analysis. Additionally, seven papers recommended by experts were manually included in the survey. Thus, a total of 126 papers were analyzed, resulting on the characterization of the research topic. Based on this study, it was possible to discover the different ways that scalability is perceived and handled in the i* community, as well as different mechanisms that have been proposed in order to tackle scalability issues.

We presented 10 concepts related to scalability described in the selected papers (Table 3). The contribution types were classified in: metamodels and formalisms; modeling processes; visual constructors; software and algorithms. The respective quantities of each contribution type are presented in Section 4.

The existing judgments about the scalability of i* models were classified by the following scale: there is no judgment, scalability is not well-treated (argumentation); scalability is not well treated (reference or evidence); scalability is well-treated (argumentation) and scalability is well treated (reference or evidence). A majority of the papers that present some judgment (67 out of 75) reported that the scalability of i* models is not well handled. A overview of some of the mechanisms that have already been proposed in order to tackle this issue is presented in Section 5.1.

Despite the fact that there are several papers on the topic, there are still many open issues to be addressed in order to improve the scalability of the i* framework. The ten challenges cited in this study (Table 4) are useful to identify emerging trends and provide an overall view of the problems tackled in the literature. In future work, we expect to establish mechanisms for performing objective comparisons of the different approaches that aim to handle i* models scalability.

In order to improve future research on the topic, the following actions are suggested: make large and complex models publicly available; interact with the industry in order to create and publish models of real systems; make the resources of scalability experiments publicly available; lastly, define metrics and exemplars for comparing different approaches.

Acknowledgements

The following Brazilian institutions have supported this work: FACEPE, CAPES and CNPQ.

References


[16] Horkoff, J.; Tong Li; Feng-Lin Li; Salnitr, M.; Cardoso, E.; Giorgini, P.; Mylopoulos, J.; Pimentel, J. Taking goal models downstream: A systematic roadmap. In: Eighth International Conference on Research Challenges in Information Science (RCIS), 2014, pp.1-12, 2014. DOI: 10.1109/RCIS.2014.6861036


Appendix - Systematic Mapping study References


Liaskos, S., Junghblut, M., Mylopoulos, J. From goal models to three-layer web-based systems: An exploratory study. In: First International Workshop on the Web and Requirements Engineering (WeRE), 2010, pp. 1–4. DOI: 10.1109/WERE.2010.5624002


Lockerbie, J., Karlsen, K., Puccio, M., Morreale, V., Bonura, S. Using requirements to define services for service-centric food traceability information systems. In: International Workshop on Service-Oriented Computing: Consequences for Engineering Requirements, SOCCER ’08, pp. 15–23, 2008. DOI: 10.1109/SOCCER.2008.8


