Tackling Large Qualitative Spatial Network of Scale-Free-Like Structure

Michael Sioutis and Jean-François Condotta

Université Lille-Nord de France, Université d'Artois
CRIL-CNRS UMR 8188
Lens, France
{sioutis,condotta}@cril.fr

Abstract. We improve the state-of-the-art method for checking the consistency of large qualitative spatial networks that appear in the Web of Data by exploiting the scale-free-like structure observed in their underlying graphs. We propose an implementation scheme that triangulates the underlying graphs of the input networks and uses a hash table based adjacency list to efficiently represent and reason with them. We generate random scale-free-like qualitative spatial networks using the Barabási-Albert (BA) model with a preferential attachment mechanism. We test our approach on the already existing random datasets that have been extensively used in the literature for evaluating the performance of qualitative spatial reasoners, our own generated random scale-free-like spatial networks, and real spatial datasets that have been made available as Linked Data. The analysis and experimental evaluation of our method presents significant improvements over the state-of-the-art approach, and establishes our implementation as the only possible solution to date to reason with large scale-free-like qualitative spatial networks efficiently.

1 Introduction

Spatial reasoning is a major field of study in Artificial Intelligence; particularly in Knowledge Representation. This field has gained a lot of attention during the last few years as it extends to a plethora of areas and domains that include, but are not limited to, ambient intelligence, dynamic GIS, cognitive robotics, spatiotemporal design, and reasoning and querying with semantic geospatial query languages [15,18,22]. In this context, an emphasis has been made on qualitative spatial reasoning which relies on qualitative abstractions of spatial aspects of the common-sense background knowledge, on which our human perspective on the physical reality is based. The concise expressiveness of the qualitative approach provides a promising framework that further boosts research and applications in the aforementioned areas and domains.

The Region Connection Calculus (RCC) is the dominant Artificial Intelligence approach for representing and reasoning about topological relations [23]. RCC can be used to describe regions that are non-empty regular subsets of some topological space by stating their topological relations to each other. RCC-8 is

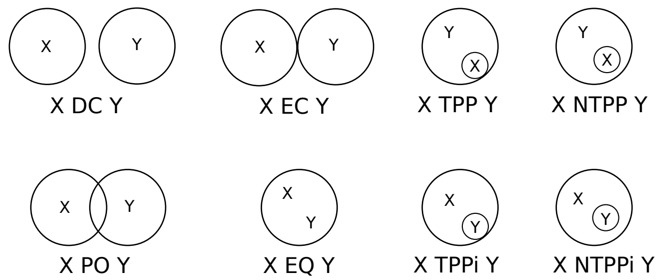


Fig. 1: Two dimensional examples for the eight base relations of RCC-8

the constraint language formed by the following 8 binary topological relations of RCC: disconnected (DC), externally connected (EC), equal (EQ), partially overlapping (PO), tangential proper part (TPP), tangential proper part inverse ($TPPI$), non-tangential proper part ($NTPP$), and non-tangential proper part inverse ($NTPPI$). These eight relations are depicted in Figure 1 ($2D$ case).

In the literature of qualitative spatial reasoning there has been a severe lack of datasets for experimental evaluation of the reasoners involved. In most cases, datasets consist of randomly generated regural networks that have a uniform node degree distribution [27] and scale up to a few hundred nodes in experimental evaluations [12, 32]. These networks are often very hard to solve instances that do not correspond to real case scenarios [31] and are mainly used to test the efficiency of different algorithm and heuristic implementations. There has been hardly any investigation or exploitage of the stuctural properties of the networks' underlying graphs. In the case where the datasets are real, they are mainly small and for proof of concept purposes, such as the one used in [9], with the exception of a real large scale and *successfully* used dataset employed in [29], viz., the admingeo dataset [14]. In fact, we will make use of this dataset again in this paper, along with an even bigger one that scales up to nearly a million of topological relations.

It has come to our attention that the real case scenario datasets we are particularly interested in correspond to graphs with a scale-free-like structure, i.e., the degree distribution of the graphs follows a power law. Scale-free graphs seem to match real world applications well and are widely observed in natural and human-made systems, including the Internet, the World Wide Web, and the Semantic Web [3, 4, 16, 30]. We argue that the case of scale-free graphs applies also to qualitative spatial networks and we stress on the importance of being able to efficiently reason with such scale-free-like networks for the following two main reasons:

- The natural approach for describing topological relations inevitably leads to the creation of graphs that exhibit hubs for particular objects which are cited more than others due to various reasons, such as size, significance, and importance. These hubs are in fact the most notable characteristic in scale-free graphs [4,16]. For example, if we were to describe the topological relations in Greece, Greece would be our major hub that would relate topologically to all of its regions and cities, followed by smaller hubs that would capture topological relations within the premises of a city or a neighborhood. It would not really make sense to specify that the porch of a house is located inside Greece, when it is already encoded that the house is located inside a city of Greece. Such natural and human-made systems are most often described by scale-free graphs [3, 4, 16, 30].

- Real qualitative spatial datasets that are known today, such as admingeo [14] and gadm-rdf[1], come from the Semantic Web, also called Web of Data, which is argued to be scale-free [30]. Further, more real datasets are to be offered by the Semantic Web community since RCC-8 has already been adopted by GeoSPARQL [22], and there is an ever increasing interest in coupling qualititave spatial reasoning techniques with linked geospatial data that are constantly being made available [19, 21]. Thus, there is a real need for scalable implementations of constraint network algorithms for qualitative and quantitative spatial constraints as RDF stores supporting linked geospatial data are expected to scale to billions of triples [19, 21].

In this paper, we concentrate on the consistency checking problem of large scale-free-like qualitative spatial networks and make the following contributions: (*i*) we explore and take advantage of the stuctural properties of the considered networks and propose an implementation scheme that triangulates their underlying graphs to retain their sparseness and uses a hash table based adjacency list to efficiently represent and reason with them, (*ii*) we make the case for a new series of random datasets, viz., *large* random scale-free-like RCC-8 networks, that can be of great use and value to the qualitative reasoning community, and (*iii*) we experimentally evaluate our approach against the state-of-the-art reasoners GQR [12], Renz's solver [27], and two of our own implementations which we briefly present here, viz., Phalanx and Phalanx$\triangledown$, and show that it significantly advances the state-of-the-art approach.

The organization of this paper is as follows. Section 2 formally introduces the RCC-8 constraint language, chordal graphs along with the triangulation procedure, and scale-free graphs and the model we follow to create them. In Section 3 we overview the state-of-the-art techniques and present our approach. In Section 4 we experimentally evaluate our approach against the state-of-the-art reasoners, and, finally, in Section 5 we conclude and give directions for future work.

We assume that the reader is familiar with the concepts of constraint networks and their corresponding constraint graphs that are not defined explicitly in this paper due to space constraints. Also, in what follows, we will refer to undirected graphs simply as graphs.

2 Preliminaries

In this section we formally introduce the RCC-8 constraint language, chordal graphs along with the triangulation procedure, and scale-free graphs together with the Barabási-Albert (BA) model.

[1] `http://gadm.geovocab.org/`

2.1 The RCC-8 constraint language

A (binary) qualitative temporal or spatial constraint language [25] is based on a finite set B of *jointly exhaustive and pairwise disjoint* (JEPD) relations defined on a domain D, called the set of base relations. The set of base relations B of a particular qualitative constraint language can be used to represent definite knowledge between any two entities with respect to the given level of granularity. B contains the identity relation Id, and is closed under the converse operation $(^{-1})$. Indefinite knowledge can be specified by unions of possible base relations, and is represented by the set containing them. Hence, 2^{B} represents the total set of relations. 2^{B} is equipped with the usual set-theoretic operations (union and intersection), the converse operation, and the weak composition operation. The converse of a relation is the union of the converses of its base relations. The weak composition $\diamond$ of two relations s and t for a set of base relations B is defined as the strongest relation $r \in 2^{\mathsf{B}}$ which contains $s \circ t$, or formally, $s \diamond t = \{b \in \mathsf{B} \mid b \cap (s \circ t) \neq \emptyset\}$, where $s \circ t = \{(x, y) \mid \exists z : (x, z) \in s \wedge (z, y) \in t\}$ is the relational composition [25,28]. In the case of the qualitative spatial constraint language RCC-8 [23], as already mentioned in Section 1, the set of base relations is the set $\{DC, EC, PO, TPP, NTPP, TPPI, NTPPI, EQ\}$, with EQ being the identity relation (Figure 1).

Definition 1. *An* RCC*-8 network comprises a pair* (V, C) *where* V *is a non empty finite set of variables and* C *is a mapping that associates a relation* $C(v, v') \in 2^{\mathsf{B}}$ *to each pair* (v, v') *of* $V \times V$*.* C *is such that* $C(v, v) \subseteq \{EQ\}$ *and* $C(v, v') = (C(v', v))^{-1}$*.*

In what follows, $C(v_i, v_j)$ will be also denoted by C_{ij}. Checking the consistency of a RCC-8 network is $\mathcal{NP}$-hard in general [26]. However, there exist large maximal tractable subsets of RCC-8 which can be used to make reasoning much more efficient even in the general $\mathcal{NP}$-hard case. These maximal tractable subsets of RCC-8 are the sets $\hat{\mathcal{H}}_8, \mathcal{C}_8$, and $\mathcal{Q}_8$ [24]. Consistency checking is then realised by a path consistency algorithm that iteratively performs the following operation until a fixed point $\overline{C}$ is reached: $\forall i, j, k,\ C_{ij} \leftarrow C_{ij} \cap (C_{ik} \diamond C_{kj})$, where variables i, k, j form triangles that belong either to a completion [27] or a chordal completion [29] of the underlying graph of the input network. Within the operation, weak composition of relations is aided by the weak composition table for RCC-8 [20]. If $C_{ij} = \emptyset$ for a pair (i, j) then C is inconsistent, otherwise $\overline{C}$ is *path consistent*. If the relations of the input RCC-8 network belong to some tractable subset of relations, path consistency implies consistency, otherwise a backtracking algorithm decomposes the initial relations into subrelations belonging to some tractable subset of relations spawning a branching search tree [28].

2.2 Chordal graphs and Triangulation

We begin by introducing the definition of a chordal graph. The interested reader may find more results regarding chordal graphs, and graph theory in general, in [13].

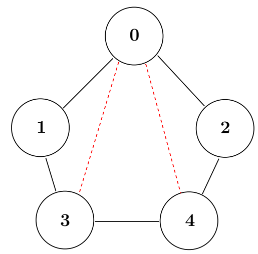


Fig. 2: Example of a chordal graph

Definition 2. *Let $G = (V, E)$ be an undirected graph. G is* chordal *or* triangulated *if every cycle of length greater than* 3 *has a chord, which is an edge connecting two non-adjacent nodes of the cycle.*

The graph shown in Figure 2 consists of a cycle which is formed by five solid edges and two dashed edges that correspond to its chords. As for this part, the graph is chordal. However, removing one dashed edge would result in a non-chordal graph. Indeed, the other dashed edge with three solid edges would form a cycle of length four with no chords. Chordality checking can be done in (linear) $O(|V| + |E|)$ time for a given graph $G = (V, E)$ with the maximum cardinality search algorithm which also constructs an elimination ordering α as a byproduct [5]. If a graph is not chordal, it can be made so by the addition of a set of new edges, called *fill edges*. This process is usually called *triangulation* of a given graph $G = (V, E)$ and can run as fast as in $O(|V| + (|E \bigcup F(\alpha)|))$ time, where $F(\alpha)$ is the set of fill edges that result by following the elimination ordering α, eliminating the nodes one by one, and connecting all nodes in the neighborhood of each eliminated node, thus, making it simplicial in the elimination graph. If the graph is already chordal, following the elimination ordering α means that no fill edges are added, i.e., α is actually a *perfect elimination ordering* [13]. For example, a perfect elimination ordering for the chordal graph shown in Figure 2 would be the ordering $1 \rightarrow 3 \rightarrow 4 \rightarrow 2 \rightarrow 0$ of its set of nodes. In general, it is desirable to achieve chordality with as few fill edges as possible. However, obtaining an optimum graph triangulation is known to be $\mathcal{NP}$-hard [5]. In a RCC-8 network fill edges correspond to universal relations, i.e., non-restrictive relations that contain all base relations.

Chordal graphs become relevant in the context of qualitative spatial reasoning due to the following result obtained in [29] that states that path consistency enforced on the underlying chordal graph of an input network can yield consistency of the input network:

Proposition 1. *For a given* RCC-*8 network $N = (V, C)$ with relations from the maximal tractable subsets $\hat{\mathcal{H}}_8, \mathcal{C}_8$, and $\mathcal{Q}_8$ and for $G = (V, E)$ its underlying chordal graph, if $\forall (i, j), (i, k), (j, k) \in E$ we have that $C_{ij} \subseteq C_{ik} \diamond C_{kj}$, then N is consistent.*

Triangulations work particularly well on sparse graphs with clustering properties, such as scale-free graphs. We are about to experimentally verify this in Section 4.

2.3 Scale-free graphs

We provide the following simple definition of a scale-free graph and elaborate on the details:

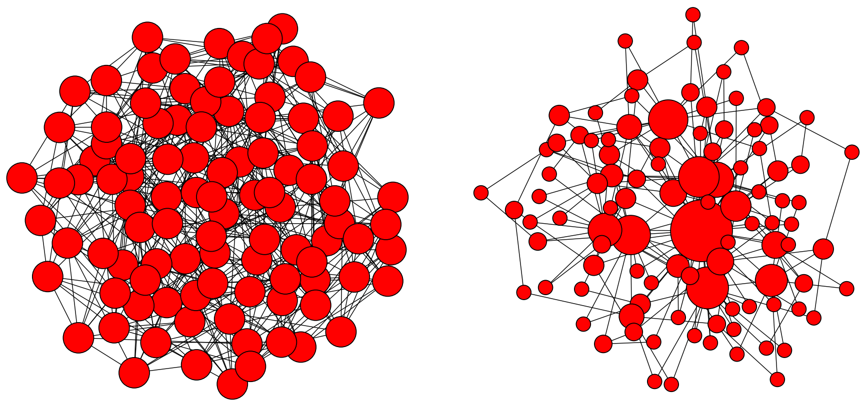
Definition 3. *Graphs with a power law tail in their node degree distribution are called* scale-free graphs *[3].*

Scale-free graphs are graphs that have a power law node degree distribution. The degree of a node in a graph is the number of connections it has to other nodes (or the number of links adjacent to it) and the degree distribution $P(k)$ is the probability distribution of these degrees over the whole graph, i.e, $P(k)$ is defined to be the fraction of nodes in the network with degree k. Thus, if there are n number of nodes in total in a graph and n_k of them have degree k, we have that $P(k) = n_k/n$. For scale-free graphs the degree distribution $P(k)$ follows a power law which can be expressed mathematically as $P(k) \sim k^{-\gamma}$, where $2 < \gamma < 3$, although γ can lie marginally outside these bounds.

There are several models to create random scale-free graphs that rely on *growth* and *preferential attachment* [7]. Growth denotes the increase in the number of nodes in the graph over time. Preferential attachment refers to the fact that new nodes tend to connect to existing nodes of large degree and, thus, means that the more connected a node is, the more likely it is to receive new links. In real case scenarios, nodes with a higher degree have stronger ability to grab links added to the network. In a topological perspective, if we consider the example of Greece that we described in Section 1, Greece would be the higher degree node that would relate topologically to new regions (e.g., Imbros) in a deterministic and natural manner. In mathematical terms, preferential attachment means that the probability that a existing node i with degree k_i acquires a link with a new node is $p(k_i) = \frac{k_i}{\sum_i k_i}$.

Among the different models to create random scale-free graphs, the Barabási-Albert (BA) model is the most well-studied and widely known one [1, 3]. The BA model considers growth and preferential attachment as follows. Regarding growth, it starts with an initial number m_0 of connected nodes and at each following step it adds a new node with $m \leq m_0$ edges that link the new node to m different existing nodes in the graph. When choosing the m different existing nodes to which the new node is linked, the BA model assumes that the probability p that the new node will be connected to node i depends on the degree k_i of node i with a value given by the expression $p \sim \frac{k_i}{\sum_i k_i}$, which is the preferential attachment that me mentioned earlier. The degree distribution resulting from the BA model is a power law of the form $P(k) \sim k^{-3}$, thus, it is able to create a subset of the total scale-free graphs that are characterised by a value γ such that $2 < \gamma < 3$. The scaling exponent is independent of m, the only parameter in the model (other than the total size of the graph one would like to obtain of course).

Scale-free graphs are particularly attractive for our approach because they have the following characteristics: (i) scale-free graphs are very sparse [10, 30],



(a) regural graph (b) scale-free graph

Fig. 3: Structures of a random regular graph with an average degree $k = 9$ and a scale-free graph with a preferential attachment $m = 2$, both having 100 nodes

and (*ii*) scale-free graphs have a clustering coefficient distribution that also follows a power law, which implies that many low-degree nodes are clustered together forming very dense subgraphs that are connected to each other through major hubs [11].

Due to the aforementioned characteristics, scale-free graphs present themselves as excellent candidates for triangulation, as sparseness keeps time complexity for triangulation low and chordal graphs also exhibit a clustering structure [13], thus, they are able to fit scale-free graphs quite effectively. As an illustration of scale-free graphs, Figure 3 depicts a random regural graph, such as the ones used for experimental evaluation in the field of qualitative spatial reasoning, and a random scale-free graph generated using the BA model. Notice that the bigger the node is, the higher its degree is (these nodes are the hubs).

3 Overview of our approach

In this section we describe our own implementations of generic qualitative reasoners that build on state-of-the-art techniques, and our practical approach of choice for tackling large scale-free-like RCC-8 networks.

State-of-the-art techniques. We have implemented Phalanx and Phalanx$\triangledown$ in Python that are the generalised and code refactored versions of PyRCC8 and PyRCC8$\triangledown$ respectively, originally presented in [29]. Phalanx$\triangledown$ is essentially Phalanx with a different path consistency implementation that allows reasoning over chordal completions of the input qualitative networks, as described in [29]. In short, Phalanx and Phalanx$\triangledown$ support small arbitrary binary constraint calculi developed for spatial and temporal reasoning, such as RCC-8 and Allen's interval algebra (IA) [2], in a way similar to GQR [12]. Further, Phalanx and Phalanx$\triangledown$

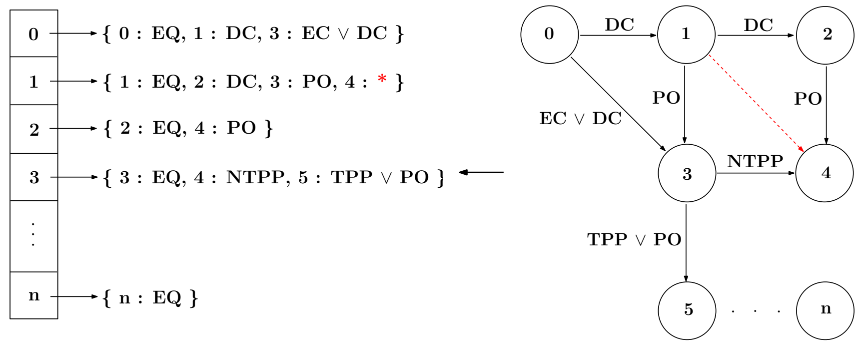


Fig. 4: Hash table based adjacency list for representing an RCC-8 network

present significant improvements over PyRCC8 and PyRCC8▽ regarding scalability and speed. In particular, the new reasoners handle the constraint matrix that represents a qualitative network more efficiently during backtracking search, i.e., they do not create a copy of the matrix at each forward step of the backtracking algorithm (as is the case with Renz's solver [27]), but they only keep track of the values that are altered at each forward step to be able to reconstruct the matrix in the case of backtracking. This mechanism is also used to keep track of unassigned variables (i.e., relations that do not belong to tractable subsets of relations and are decomposed to subrelations at each forward step of the backtracking algorithm) that may dynamically change in number due to the appliance of path consistency at each forward step of the backtracking algorithm. For example, the path consistency algorithm can prune a relation that belongs to a tractable subset of relations into an untractable relation, and vice versa. This allows us to apply the heuristics that deal with the selection of the next unassigned variable faster, as we keep our set of unassigned variables minimal. The path consistency algorithm implementation has been also modified to better handle the cases where weak composition of relations leads to the universal relation. In these cases we can continue the iterative operation of the path consistency algorithm since the intersection of the universal relation with any other relation leaves the latter relation intact. Finally, there is also a weight over learned weights dynamic heuristic for variable selection, but we still lack the functionality of restart and nogood recording that has been implemented in the latest version of GQR, release 1500[2], in the time of writing this paper [33]. In any case, and in defense of GQR which was found to perform poorly in [29] under release 1418, we state that the latest version of GQR has undergone massive scalability improvements and is currently the most complete and fastest reasoner for handling reasonably scaled random *regular* qualitative networks. However, in the experimens to follow, we greatly outperform GQR for large QCNs of scale-free-like structure. At this point, we can also claim that Renz's solver [27] has been fairly outdated, as it will become apparent in the experiments that we employ it for.

[2] `http://sfbtr8.informatik.uni-freiburg.de/R4LogoSpace/downloads/gqr-1500.tar.bz2`

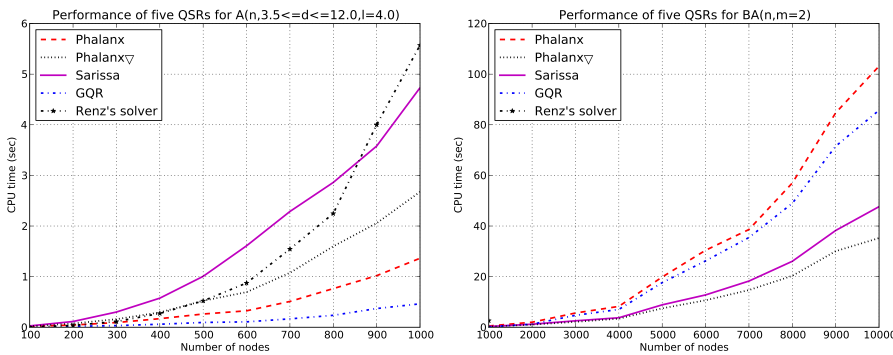
We improve the state-of-the-art techniques for tackling large scale-free-like RCC-8 networks by opting for a hash table based adjacency list to represent and reason with the chordal completion of the input network. The variables of the input network (or the nodes) are represented by index numbers of a list, and each variable (or node) is associated with a hash table that stores key-value pairs of variables and relations. Figure 4 shows how an example RCC-8 network is represented by our hash table based adjacency list approach. Self-loops (of the identity relation EQ) have been omitted from the network. The dashed edge $(1, 4)$ corresponds to a fill edge that results after triangulating the initial non-chordal network consisting of solid edges. This fill edge is stored in the hash table based adjacency list as a universal relation, denoted by symbol $*$. For a given RCC-8 network $N = (V, C)$ and for $G = (V, E)$ its underlying chordal graph, our approach requires $O(|V|+|E|\cdot b)$ memory, where b is the size needed to represent a relation from the set of relations 2^{B} of RCC-8. Having a constraint matrix to represent the input RCC-8 network (that is typically used by the state-of-the-art reasoners), results in a $O(|V|^2 \cdot b)$ memory requirement, even if chordal graphs are used leaving a big part of the matrix empty, as is the case with Phalanx$\bigtriangledown$, or Sparrow for IA [8]. Further, we still retain an $O(1)$ average access and update time complexity which becomes $O(\delta)$ in the amortized worst case, where δ is the average degree of the chordal graph that corresponds to the input network. Given that we target large scale-free-like, and, thus, sparse networks, this only incures a small penalty for the related experiments performed. The path consistency implementation also benefits from this approach as the queue data structure which is based on has to use only $O(|E|)$ of memory to store the relations compared to the $O(|V|^2)$ memory requirement of Phalanx, GQR, and Renz's solver. Regarding triangulation, our hash table based adjacency list is coupled with the implementation of the maximum cardinality search algorithm and a fast fill in procedure (as discussed in Section 2.2), as opposed to the heuristic based, but rather naive, triangulation procedure implemented in [29]. Though the maximum cardinality search algorithm does not yield minimal triangulations if the underlying graph of the input network is not chordal, it does guarantee than no fill edges are inserted if the graph is indeed chordal. In addition, even for the non-chordal cases we obtain much better results with this approach and have a fine trade-off between time efficiency and good triangulations.

These techniques are implemented under the hood of our new reasoner which is called Sarissa. Sarissa, as with our other tools presented here, is a generic and open source qualitative reasoner written in Python.[3]

4 Experimental evaluation

In this section we compare the performance of Sarissa with that of Renz's solver [27], GQR (release 1500) [12], Phalanx, and Phalanx$\bigtriangledown$, with their best performing heuristics enabled.

[3]All tools and datasets used here can be acquired upon request from the authors or found online in the following address: `http://www.cril.fr/~sioutis/work.php`.



(a) performance for regural networks (b) performance for scale-free-like networks

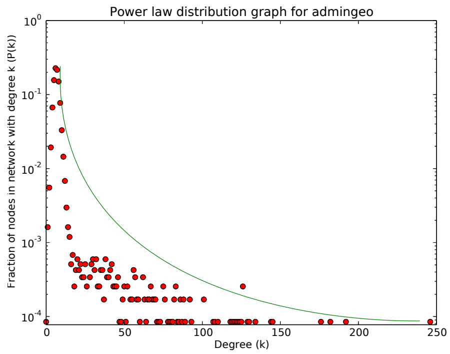
Fig. 5: Performance of five reasoners for randomly generated networks

We considered both random and real datasets. Random datasets consist of RCC-8 networks generated by the usual A(n, d, l) model [27] and *large* RCC-8 networks generated by the BA(n, m) model which we first introduce for benchmarking qualitative spatial reasoners in this paper. In short, model A(n, d, l) creates random regural networks (like the one depicted in Figure 3a) of size n, degree d, and an average number l of RCC-8 relations per edge, whereas model BA(n, m) creates random scale-free-like networks (like the one depicted in Figure 3b) of size n and a preferential attachment value m. For model BA(n, m) the average number of RCC-8 relations per edge defaults to $|B|/2$, where B is the set of base relations of RCC-8. Real datasets consist of admingeo [14] and gadm-rdf[4] that comprise 11761/77907 nodes/edges and 276728/590865 nodes/edges respectively. In short, admingeo describes the administrative geography of Great Britain using RCC-8 relations, and gadm-rdf the world's administrative areas likewise. The experiments were carried out on a computer with an Intel Core 2 Quad Q9400 processor with a CPU frequency of 2.66 GHz, 8 GB RAM, and the Precise Pangolin x86_64 OS (Ubuntu Linux). Renz's solver and GQR were compiled with gcc/g++ 4.6.3. Sarissa, Phalanx, and Phalanx$\triangledown$ were run with PyPy[5] 1.9, which implements Python 2. Only one of the CPU cores was used for the experiments.

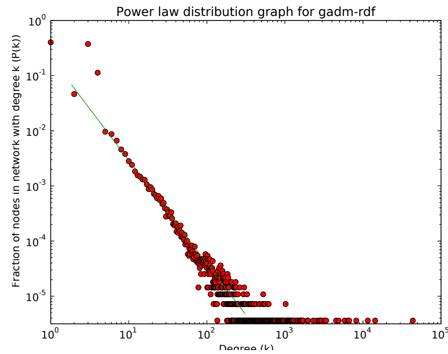
Random datasets. For model A(n, d, l) we considered network sizes between 100 and 1000 with a 100 step and $l = 4$ $(= |B|/2)$ relations per edge. For each size series we created 270 networks that span over a degree d between 3.5 and 12.0 with a 0.5 step, i.e., 15 network instances were generated for each degree. The results are shown in Figure 5a. GQR clearly outperforms all other reasoners with Phalanx coming close 2^{nd} and Renz's solver last. In the particular case of Sarissa and Phalanx$\triangledown$ that use chordal graphs we note that they pay an extra cost for calculating the triangles of relations for each appliance of path consistency as these are not precomputated and stored in advance for memory efficiency [29].

[4] http://gadm.geovocab.org/

[5] http://pypy.org/

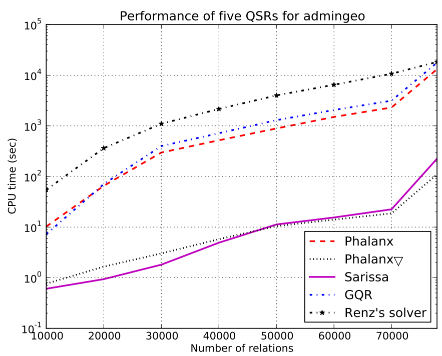


(a) degree distribution for admingeo

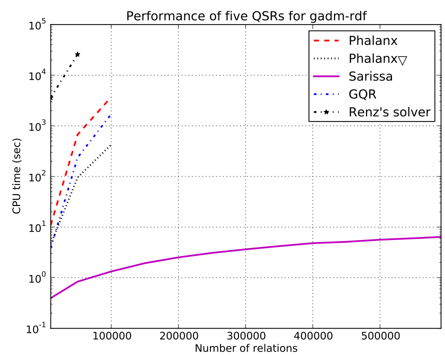


(b) degree distribution for gadm-rdf

Fig. 6: The figure provides evidence of the power law node degree distribution of the real datasets considered



(a) performance for admingeo



(b) performance for gadm-rdf

Fig. 7: Performance of five reasoners for real datasets

Sarissa also pays an additional cost for not being able to access or update relational values in constant worst case time as it does not use a matrix. It is a fact that random regural networks are not triangulated very efficiently with our approach, which results in dense chordal graphs in most of the cases.

For model $\mathsf{BA}(\mathsf{n}, \mathsf{m})$ we considered 30 networks for each size between 1000 and 10000 with a 1000 step and a preferential attachment value of $m = 2$. We found that for this specific value of m and for the network sizes considered, the networks lie within the *phase transition* region, where it is equally possible for networks to be consistent or inconsistent, thus, they are harder to solve. The results are shown in Figure 5b. Sarissa and Phalanx$\triangledown$ outperform all other reasoners by a large scale, and Renz's solver was able to solve only the networks of 1000 nodes (slower than all others) as it quickly hit the memory limit of our computer due to many recursive calls (leading to storing many copies of the matrix). We note that Sarissa is still burdened with the additional cost of not being able to access or update relational values in constant worst case time. To the best of our knowledge, the random datasets used in this paper are the biggest ones to date of all others that exist in literature.

Real datasets. For experimenting with our real datasets, viz., admingeo and gadm-rdf, we created constraint networks of different size, by taking into account a small number of relations from the initial dataset and increasing it at every next step. Of course, for both datasets, the whole dataset was used as a final step. To backup our argument about the scale-free-like structure of real RCC-8 networks, we present Figure 6 that displays the power law degree distribution of our real networks. As gadm-rdf is a very large network, we display its degree distribution in log-log scale where the power law function is seen as a straight line [3]. The results for admingeo are shown in Figure 7a. Sarissa and Phalanx$\bigtriangledown$ have approximately equal performance and significantly outperform all other reasoners. In the final step, both reasoners run in $\sim$ 150 sec, when the 3^{rd} best reasoner for this experiment, viz., Phalanx, runs in $\sim$ 4 hours. Up to this point we have considered networks that fit the size of the matrix that accompanies all state-of-the-art reasoners. We proceed with gadm-rdf, a dataset almost 30 times bigger than admingeo. The results for gadm-rdf are shown in Figure 7b. Sarissa is the only implementation that was able to reason with the whole dataset. Sarissa completes the final step of the 590865 relations in under $\sim$ 7 sec, when the 2^{nd} best reasoner for this experiment, viz., Phalanx$\bigtriangledown$, can only reason up to 100000 relations in $\sim$ 7 min. GQR reasons up to 100000 relations in $\sim$ 28 min, Phalanx in double the time of GQR, and, finally, Renz's solver reasons up to 50000[6] relations in $\sim$ 7 hours. Gadm-rdf is the biggest real dataset to date to have been succesfully employed in an experimental evaluation of qualitative spatial reasoners. Surprisingly, Sarissa runs the gadm-rdf experiment faster than the admingeo one, but this is due to more relations being inferred in the latter case as a result of dataset particularities that affect the reasoning process.

At this point we conlude our experimental evaluations. Due to space constraints we omitted several graphs that would display the amount of edges considered by each implementation, the effect of the triangulations, and experiments with the $\mathcal{NP}_8$ class of RCC-8 relations [27].

5 Conclusion and Future work

In this paper we have presented an approach that employs chordal graphs and a hash table based adjacency list implementation to tackle large scale-free-like qualitative spatial networks, and goes well beyond the state-of-the-art qualitative spatial reasoners which were found to come short of the task.

One could argue that even though being able to tackle a real dataset of nearly a million relations fairly easily, our approach is still far from the billion relations goal set in [19, 21]. However, for the case of tractable RCC-8 networks

[6] In practice, Renz's solver was able to fit the 100000 relations of the next step, but judging by its overall performance it would require several days to reason with them.

and due to a particular patchwork property presented in [17], our approach allows for building a consistent RCC-8 enriched database incrementally. This can be achieved by initially reasoning with a small part of the dataset, and then for every new piece of data only considering relevant existing RCC-8 relations (if any), reasoning with the resulted fused piece of information, and continuing the process, while maintaining chordality [6]. In a likewise manner, for quering or updating data one would only have to consider relations relevant to his regural or update query. Future work consists of exploring this solution for datasets that scale up to billions of relations, and also further exploring and optimizing on the hierarchical structure that real datasets present, as argued in [21]. In particular, it would be interesting to explore which relations are used more than others in real datasets and whether this could be of some use or not. We would also like to investigate if stuctural or hierarchical properties are observed in real IA networks which we are currently in the process of obtaining from temporally enriched datasets. Another research direction would be to explore if SAT encodings are able to tackle large scale-free-like RCC-8 networks, although given that SAT encodings become too large when network sizes increase beyond a few hundred nodes this would be highly unlikely.

Finally, we feel that it is very important to motivate the qualitative reasoning community to get involved with the structure that real datasets present, and to this direction large random scale-free-like networks can be of great use and value to further improve existing reasoners and present (possibly mixed) solutions that can scale up to billions of relations.

Acknowledgments. This work was funded by a PhD grant from Université d'Artois and region Nord-Pas-de-Calais.

References

1. Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47–97 (2002)
2. Allen, J.F.: Maintaining knowledge about temporal intervals. CACM 26, 832–843 (1983)
3. Barabasi, A.L., Albert, R.: Emergence of scaling in random networks. Science (New York, N.Y.) 286, 509–512 (1999)
4. Barabasi, A.L., Bonabeau, E.: Scale-Free Networks. Scientific American pp. 50–59 (2003)
5. Berry, A., Blair, J.R.S., Heggernes, P.: Maximum Cardinality Search for Computing Minimal Triangulations. In: WG (2002)
6. Berry, A., Heggernes, P., Villanger, Y.: A vertex incremental approach for maintaining chordality. Discrete Mathematics 306 (2006)
7. Bollobás, B.: Mathematical results on scale-free random graphs. In: Handbook of Graphs and Networks. pp. 1–37. Wiley (2003)
8. Chmeiss, A., Condotta, J.F.: Consistency of Triangulated Temporal Qualitative Constraint Networks. In: ICTAI (2011)
9. Christodoulou, G., Petrakis, E.G.M., Batsakis, S.: Qualitative Spatial Reasoning Using Topological and Directional Information in OWL. In: ICTAI (2012)

10. Del Genio, C.I., Gross, T., Bassler, K.E.: All Scale-Free Networks Are Sparse. Phys. Rev. Lett. 107, 178701 (2011)
11. Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.F.F.: Pseudofractal scale-free web. Physical Review E 65, 066122+ (2002)
12. Gantner, Z., Westphal, M., Wölfl, S.: GQR-A Fast Reasoner for Binary Qualitative Constraint Calculi. In: AAAI Workshop on Spatial and Temporal Reasoning (2008)
13. Golumbic, M.C.: Algorithmic Graph Theory and Perfect Graphs. Elsevier Science, 2nd edn. (2004)
14. Goodwin, J., Dolbear, C., Hart, G.: Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web. TGIS 12, 19–30 (2008)
15. Hazarika, S.: Qualitative Spatio-Temporal Representation and Reasoning: Trends and Future Directions. Igi Global (2012)
16. Hein, O., Schwind, M., Knig, W.: Scale-Free Networks - The Impact of Fat Tailed Degree Distribution on Diffusion and Communication Processes. Wirtschaftsinformatik 47, 21–28 (2006)
17. Huang, J.: Compactness and its implications for qualitative spatial and temporal reasoning. In: KR (2012)
18. Koubarakis, M., Kyzirakos, K.: Modeling and Querying Metadata in the Semantic Sensor Web: The Model stRDF and the Query Language stSPARQL. In: ESWC (2010)
19. Koubarakis, M., Kyzirakos, K., Karpathiotakis, M., Nikolaou, C., Sioutis, M., Vassos, S., Michail, D., Herekakis, T., Kontoes, C., Papoutsis, I.: Challenges for Qualitative Spatial Reasoning in Linked Geospatial Data. In: BASR (2011)
20. Li, S., Ying, M.: Region connection calculus: Its models and composition table. Artif. Intell. 145, 121–146 (2003)
21. Nikolaou, C., Koubarakis, M.: Querying Incomplete Geospatial Information in RDF. In: SSTD (2013)
22. Open Geospatial Consortium: OGC GeoSPARQL - A geographic query language for RDF data. OGC® Implementation Standard (2012)
23. Randell, D.A., Cui, Z., Cohn, A.: A Spatial Logic Based on Regions and Connection. In: KR (1992)
24. Renz, J.: Maximal Tractable Fragments of the Region Connection Calculus: A Complete Analysis. In: IJCAI (1999)
25. Renz, J., Ligozat, G.: Weak Composition for Qualitative Spatial and Temporal Reasoning. In: CP (2005)
26. Renz, J., Nebel, B.: Spatial Reasoning with Topological Information. In: Spatial Cognition (1998)
27. Renz, J., Nebel, B.: Efficient Methods for Qualitative Spatial Reasoning. JAIR 15, 289–318 (2001)
28. Renz, J., Nebel, B.: Qualitative Spatial Reasoning Using Constraint Calculi. In: Handbook of Spatial Logics, pp. 161–215 (2007)
29. Sioutis, M., Koubarakis, M.: Consistency of Chordal RCC-8 Networks. In: ICTAI (2012)
30. Steyvers, M., Tenenbaum, J.B.: The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. Cognitive Science 29, 41–78 (2005)
31. Walsh, T.: Search on High Degree Graphs. In: IJCAI (2001)
32. Westphal, M., Wölfl, S.: Qualitative CSP, Finite CSP, and SAT: Comparing Methods for Qualitative Constraint-based Reasoning. In: IJCAI (2009)
33. Westphal, M., Wölfl, S., Li, J.J.: Restarts and Nogood Recording in Qualitative Constraint-based Reasoning. In: ECAI (2010)