COMPLEXNETGIS: A TOOL FOR THE ANALYSIS OF COMPLEX SPATIAL NETWORKS

Simone Caschili

Abstract. ComplexNetGIS, an integrated ArcGIS® tool, was developed based on demand for a specific instrument to implement analyses on large spatial networks. ComplexNetGIS helps to integrate the analysis provided by Complex Network Theory together with tools for geographical visualization and analysis. Due to the wide application of Complex Network Theory to different fields of research, ComplexNetGIS can be useful in different areas ranging from studies in applied physics to the analysis of facilities management and transportation planning.

Keywords: Complex Networks, Complexity, Network Analysis, GIS.

1 Introduction

Over the past 50 years the interdisciplinary topic of Complexity has been used to describe and model social phenomena together with space-time features. Complexity theory is influential in a variety of scientific fields. Alan Wilson [1] conceived of an elegant and easy approach to the understanding of this topic: he first asked himself what characterized a system as complex and then which features were of particular importance for complex systems.

In order to begin answering these questions Wilson divided the systems into two categories; simple and complex systems. Simple systems are systems that can be modelled by a small number of variables where as complex systems are systems whose elements are singularly observable and easily understood; however, when dealing with complex systems, it is difficult to forecast possible future configurations due to the non linear and interdependent dynamics.

Building upon this approach, scholars of applied physics proposed a new scientific paradigm for the analysis and modelling of both natural and technological (or man made) activities around the turn of the last century. This paradigm was named Complex Network Theory (CNT). During the late 1990’s, one of the pioneer scholars,
Albert-László Barabási together with his research group at Notre Dame (Indiana University, USA) studied the properties of the World Wide Web. Barabási analyzed some Internet branches in order to characterize its topology, using a software machine similar to that used by Internet search engines such as Google. In the beginning, the research group’s hypothesis was that the World Wide Web is a random network. This hypothesis was supported by the idea that each Internet user could create a web page randomly without any predetermined rule. They expected to find a network where there was a high probability of having nodes with an average number of connections. However, the results refuted their hypothesis. Surprisingly, the Internet is a network composed of a majority of nodes with few connections and a small amount of nodes that have high connectivity (network’s hubs). Graphs that demonstrate these properties are also named Scale-free networks. In order to compare and contrast the differences between random and Scale-free networks, Barabási [2] relates the first to the American road network and the second to the international air route network respectively. Each town has at least one small airport which is occasionally linked to other small sized airports. More often, small sized airports have direct flights to big international airports, ordinarily called hubs. In contrast, even the busiest nodes (usually cities) within a road network have a number of connections of the same order of magnitude of the mean value.

Starting from applied physics studies, the Complex Network (CN) paradigm has received considerable interest in a number of fields. The exponential application of network modelling has shown that some of the properties exhibited by the World Wide Web, are also common to many other systems, often not related to one another. Social relations (i.e. the spread of virus, the author collaboration network, citation network), technological networks (Internet, World Wide Web, power grid), ecology (evolution and extinction of species) and linguistics (studies on the frequency of words in texts) surprisingly have in common the same properties found by Barabási for the World Wide Web.

Real world phenomena can also be modelled as networks since they are made of interconnected and interacting elements that influence one another. From this perspective, territory is a medium where phenomena happen as result of the spatial dynamics among people and the natural environment.

Thus, there is still a growing yet unsatisfied need to model and visualize complex networks in a spatial domain. Geographic Information Systems (GIS) provides a useful instrument to integrate the analyses and visualization of spatial networks. But, as discussed by Batty [3]:

“… getting such ideas into GIS is thus an enormous challenge for it is much easier to see GIS as providing some convenient form of visualization, data storage and manipulation technology rather than a vehicle on which to make contemporary geographical theory applicable and practicable.”

Traditional basic GIS software does not allow one to model CN features such as interactions among agents. For this reason GIS experts have created a number of ad-hoc extensions able to deal with specific network analyses.
Some GIS tools provide useful graphic interfaces and geographic databases for the management of network assets and flows [4]. Some vendors have already developed corporate GIS software to analyze facilities management, such as gas, electricity, water, telecommunications and transportation. Examples of those kinds of GIS software are ArcGIS® Network Analyst and GE Small World.

In the field of Complex Network Analysis (CNA) several studies took into account network spatial features such as airway networks [5], street networks [6], commuting networks [7, 8] and social networks [9].

When an end-user deals with spatial network analysis he or she is required to have a deep understanding in both network analysis and geography. It is unusual to find scholars with a good deal of experience in both fields since skilled network analysts usually come from an applied physics background while GIS users from geography and spatial analysis fields. Thus, it is increasingly necessary to share basic network information and, in some cases, integrate data sets in order to carry out more complex network analysis operations [10].

In this contribution we propose a tool that attempts to fill the gap between GIS users (geographers) and network analysts (physicists) when dealing with spatial networks. The tool was developed as a ArcToolBox element within the ESRI ArcGIS® software. It was developed using Python programming language together with the use of the network library NetworkX [11].

In the subsequent sections we present CNT cornerstones, illustrate the role GIS tools may play in CNA and what prompted us to develop ComplexNetGIS. In addition to this background information, we explore ComplexNetGIS functionalities and its Graphical User Interface. In the final section we present examples of possible applications and the conclusions.

2 Complex Networks and Spatial Analysis

Networks are ubiquitous in the everyday world. All networks share a common construct of nodes connected together by links. The very simple concept of one location connecting to another quickly becomes an extremely complex phenomenon as the number of nodes and connections increase. Over the last 20 years, the concept that simple networks evolve into incredibly complex and dynamic networks has produced an abundance of research in physics, computer science, molecular biology, sociology, and many other fields; all contributing to the growth of CN studies. Presently, this field (CN studies) has not sufficiently taken into account the geography of networks. The vast majority of research on CN revolves around abstract networks where geographical properties, such as the location of the nodes, are not considered.

Early graph theory analysis was confined to relatively small networks with a computationally manageable number of edges and vertices. The earliest studies had to do with geographical analysis, especially in the field of transportation networks. The studies of Kansky [12] and Haggett and Chorley [13] used the graph theoretical
implications of transportation networks to help explain aspects of regional and national economies. Milgram [14] paid attention to the Small-world problem (i.e. the power of certain networks to easily connect to apparently distant nodes) specifically in the context of geography and social psychology. Milgram tested the “six degrees of separation law” that refers to the idea that every person is six steps away from every other person on the Earth in terms of interpersonal acquaintance.

In the field of urban planning, Stoneham [15] investigated the spatial aspects of the Small-world problem.

Around the same period, Erdös and Rényi [16, 17] were working on a theoretical study focused on large complex graphs. They applied probabilistic methods to model problems regarding large graphs [2]. They studied large graphs by statistical algorithms where nodes were randomly connected. Erdös and Rényi found that when vertices were connected in this fashion their topology (the degree probability distribution—i.e. number of connection per node) was conformed to a Poisson distribution [2].

Watts and Strogatz [18] verified that some complex networks showed particular clustering patterns (nodes that are close and well connected to each other). Analyzing several large data sets, they found that many real world networks were not entirely random but instead displayed significant clustering at the local level.

Thus, summarizing the preceding concepts, systems become more complex as they accumulate interactions [3]. The degree of complexity (number of interactions between a system’s agents) grows according to power laws. More importantly, a selective growth exists. Edges link nodes selectively and optimise the system as a whole. This concept has been applied through the Barabási–Albert (B-A) model [19] where the network evolution could be epitomized by the clique “the rich gets richer”. In the B-A model the more links a node has the more links a node will get.

The CN paradigm has been also studied in the spatial domain within some of the more famous pioneering studies about the applications of CNT to geographic networks. Guimerà et al. [5] found that the worldwide air transportation network is a Small-world network in which the number of connections and the number of shortest paths have Scale-free distributions. They found that the busiest nodes are not always the most central in the network. They hypothesized that the presence of a multi-community structure could affect this behaviour and could be understood in terms of both geographical and political considerations.

Porta et al. [6] performed a comparison between primal and dual graph representations of urban street networks of the cities of Ahmedabad, Barcelona, San Francisco, Venezia, Wien and Walnut Creek. They found that Small-world properties emerge as a general rule throughout all cases.

Several studies about commuting networks [7, 8] have demonstrated that commuting networks are similar to a Small-world random network [18].

Thus, there is a growing interest in the inclusion of spatial and geographic features in CNT. Scholars have been calling for specific instruments to integrate these analyses [3]. At the same time the availability of such tools may launch a fruitful integration between geography and physics in the modelling of spatial complex systems. In the
In the introduction we pointed out that there is a need to integrate the analysis provided by CNT with tools for geographic visualisation and analysis. In this paragraph, we discuss this assumption based on the theoretical analysis of alternative approaches to integrate GIScience and CNT as carried out by De Montis et al. [20].

According to Jiang et al. [21] GIScience has paid little attention to methods which involve relations, connections and interactions. More recently, new (partially successful) models have been proposed to solve these issues. The ESRI™ geo-database network model [22] integrates a geometric network consisting of a planar network, with a logical network. This network model provides a logical structure which collects features that form a system (a collection of vertices and edges). In CNA a logical network is commonly used to model the relationships between agents (i.e. nodes in network modelling). Integration with the geometric network model allows one to take into account the spatial dimension through the coordinates of the connected vertices. Moreover the geometric network allows one to store attributes both for nodes and edges. This capability adds promising GIS modelling for the analysis of weighted complex networks. New attributes for nodes and links can indeed be calculated in a GIS environment. This process allows a user to take into account complex relationships between nodes, links, and the surrounding environment using simple (such as buffering or overlaying) or advanced spatial analysis functions. The resulting database can be used to further apply traditional network analysis (e.g. shortest path measures, near neighbours analysis, etc…).

These functionalities are integrated in the recent release of ArcGIS®. The proprietary nature of ESRI™ products does not allow users to develop and extend its functionalities, but just to customize them according to a set of predefined rules. Many of the functionalities implemented have been developed to solve network management issues but do not suit topological, cluster and flow network analyses which are the cornerstones of CNT. This is what we refer to when we mention a lack of integration between GIS and CNT. We need a tool that allows network analysts to inspect topological and dynamic features of spatial networks, extend CNA with measures and approaches characteristic of geography (or GIScience) and visualize the results on maps and charts.

Campagna et al. [7, 20] proposed four theoretical approaches for the integration of CNT in a GIS environment. Each of those approaches provides the user with a different benefit that GIS may offer to CNA. These approaches span from the representation of CNA results, to feeding CNA input datasets and performing complex network analyses [20]. According to that classification, ComplexNetGIS may be categorized into integration GISF-II, which are Network models with spatial
attributes that can be fully modelled and analysed in a GIS environment, by developing some programming codes.

ComplexNetGIS provides several network measures and indices taken from CNT and transport analysis. The next few paragraphs introduce the tool architecture and its graphical interface.

4 ComplexNetGIS: the tool architecture

ComplexNetGIS uses ArcGIS® as Graphical User Interface (GUI) while the tool’s kernel is implemented through a Python scripting method together with the use of the library NetworkX [11]. Our current implementation involves two components:

- **A conversion module**, which allows the users to build a topological network frame whether or not the input file is a network data file or a shapefile;
- **A network processing module**, which allows the users to perform network analysis. Its engine exploits the network library NetworkX [11] developed in Python programming language.

Fig. 1 provides the architecture for the ComplexNetGIS tool. The ArcGIS® GUI serves as an input data management system that maintains all the spatial data and provides support on-screen visualization. The spatial data includes line coverage representing the network edges (such as streets, pipelines, spatial interactions, etc) and point coverage representing the network vertices (such as locations of activities, street intersections, pipeline joints, agents, etc).

The tool was built with the aim to read existing network data without having to digitalize them in ArcGIS®. The conversion module pursues this goal. The user can both load a line shapefile and a ASCII file. Corporate GIS network tools on the market mostly lack the advantage of import/export standard network data, in formats such as edge list format, GML or Pajek. ComplexNetGIS is indeed able to read the more highly utilized network data structures, such as edge list and Pajek standards.

Because we are tackling spatial domains, the input network file must reference spatial location. Thus, in planar networks, node coordinates X and Y are always required while elevation coordinate Z and a forth m value are optional attributes. Each link may also supply a weight representing the level of interaction between each pair of connected nodes. The geometrical attributes X, Y, Z, m can be provided with a edge list file or .NET Pajek network standard.

Though, some network software does not even support topology or network analysis (such as Network Analyst extension by ArcGIS®) the foundations of complex network studies theory.
The core of the tool is the network processing module, a computational kernel that consists of three sub-modules:

- **Topological analysis**;
- **Weighted analysis**;
- **Cluster analysis**.

At present, the topological and weighted analysis modules are fully exploited while cluster analysis is under advanced development. Below we review the functions implemented in each of the first two modules. It is worth noting that ComplexNetGIS is an open source tool distributed under the GNU General Public License. The open source nature of the tool code will allow skilled programming users to extend the tool functionalities according to his or her needs.

### 4.1 Topological analysis module

The topological module integrates two major network analyses: the first refers to some of the classical measures used in CNA. The second offers a characterization of transportation networks, according to the indices proposed by Kansky [12]. Grouped under the measures derived from CNA, ComplexNetGIS implements: *Degree analysis, Clustering analysis, Degree centrality, Betweenness centrality, Topological Diameter, Spatial Diameter, Number of cycles, Max degree, Average degree, Min degree.*
ComplexNetGIS also develops a number of transportation indexes [24]: *Average Network Detour Index*, *Alpha index*, *Beta index*, *Gamma index*, *Eta index*, *Theta index*, *Topological Pi index*, *Spatial Pi index*.

Those indices are used in transportation planning to analyse the network efficiency in terms of graph structure. To our knowledge, no ArcGIS® tool has integrated these measures yet. Thus, ComplexNetGIS allows ArcGIS® to extend its functionalities also in GIS transportation analysis. Through ComplexNetGIS, ArcGIS® can provide measures to analyse topological efficiency of networks together with the management functionalities provided by Network Analyst extension.

### 4.2 Weighted analysis module

The connections in many networks are not merely binary entities, either present or not, but have associated weights that record their strengths relative to one another. The ComplexNetGIS conversion module allows one to add to each link a weight representative of the level of interaction between two nodes. Some of the possible measures proposed by weighted network studies are implemented in this module:

- **Strength**, the strength is the sum of the weight of the links attached to a node;
- **Disparity** as proposed by Barthelemy et al. [23], it measures the pattern of flow distribution in the network; the mathematical formulation is given by the sum of the square of each weight attached to a random node divided by the square of its strength.

At present the *Weighted analysis module* provides only these two measures. Future implementations aim at extending those functionalities that, as we have already discussed, could also be improved by users through the use of NetworkX library in python scripting language.

### 5 ComplexNetGIS: the Graphical User Interface

ComplexNetGIS is implemented as ArcGIS® Arctoolbox element. The GUI consists of modules each of which allows the functions illustrated in section 4 to be performed.

In Fig. 2 we display a snapshot of ComplexNetGIS within Arctoolbox GUI and an example of the application of this tool on the commuting network of Sardinia. The number of commuters per link is visualized using the stop/go symbols and the line width is progressively modified accordingly to the number of commuters per link.
Fig. 2. The ComplexNetGIS GUI. The commuting network of Sardinia is visualized.

The interface allows users to access the geo-processing functions that are also available using the command line. When a module is launched an input interface appears and the user may enter the required data. The input data can be either required or optional. The interface then sends the input information to the underlying script and runs it. Each input field also has a brief description on the right side of the window and in each module a help reference document is provided.

6 Conclusions

CNT lacks the inclusion of network spatial features in its theoretical background. When dealing with spatial or territorial phenomena, CNT takes into account the topological structure of networks but it does not consider the spatial location of nodes and their relative positions.

This epistemological issue may be ascribed to both the lack of adequate instruments and experts in spatial networking. In order to perform CNA in the field of geography, a trans-disciplinary effort is required to build tools able to integrate relevant information into computerized software. Geographic Information Systems appear to be a convenient platform to integrate CNA and spatial analysis. This variety of software allows one to store, manipulate, analyze and visualize both spatial and a-spatial information.

With the needs of CNA users in mind, we created the tool ComplexNetGIS. It was developed as a built-in tool within the ESRI® software ArcGIS®.

The use of the ComplexNetGIS tool may be helpful for researchers and users in many different fields as it allows one to perform analysis and visualize results within GIS software. ArcGIS® GUI is relatively user friendly and therefore makes it possible for
“newbie” GIS users to enjoy the output of their network analysis. Furthermore, expert GIS users are able to take advantage of ComplexNetGIS by implementing ArcGIS® network analysis without needing to use third-party software. Because ComplexNetGIS is under the GNU General Public License, it is accessible to a wide audience. We hope that other researchers will use the tool, tailor it to their own needs and perhaps use the information gained from it in real-world scenarios. We aim to create a community of developers interested in extending and improving functionalities and capabilities of ComplexNetGIS in the near future.

Acknowledgments. The author wishes to thank Michele Campagna and Andrea De Montis for the useful comments and discussions.

References


1 The tool is available on the web page http://people.unica.it/adm/complexnetgis/