DETECTION OF ORGANIZATIONS IN LARGE GRAPHS : APPLICATION ON RISK MANAGEMENT

Cyrille Bertelle, Antoine Dutot, Michel Nabaa[1], Damien Olivier
LITIS
Université du Havre
F-76085 Le Havre
25, rue Phillipe Lebon
{Cyrille.Bertelle; Antoine.Dutot; Michel.Nabaa; Damien.Olivier; }@univ-lehavre.fr
Pascal Mallet
CODAH
F-76085 Le Havre
19, Rue Georges Braque
Pascal.Mallet@agglo-havraise.fr

Abstract

Recent events have shown that our agglomerations are vulnerable in front of emergency situations. There are a lot of factors to consider and one of them is the structure of towns generated by road networks. One way to study these structures is to modelize the networks by graphs and to use the theory developed on them. Recently, some works have studied the detection of communities in large graphs. A community in a graph is a set of connected nodes strongly connected to each other and less connected to nodes from other communities. The aim of detection is to identify communities according to a predefined criteria but without specifying the number of communities and the size of communites. The known algorithms find communities but without taking into account the time. We propose a dynamic algorithm of communities detection and we explain how it can be applied on Le Havre agglomeration, for example, to estimate the vulnerability related to the road network use by vehicles.

1 Introduction

Modeling the population flow aims to provide a simplified representation of the population displacement, which is a complex phenomena. It must take into ac-

[1]Corresponding author. Authors are alphabetically ordered. Michel Nabaa research is supported by Haute-Normandie Region.

count the displacement of individuals in their normal daily activities or under exceptional activities requiring an evacuation in a danger case. Models must eventually allow a better understanding of the parameters influencing the flow evolution and provide decision making tools to know the spatial and temporal distribution of people in a given urban zone. The objective is therefore to realize a simulation that can be used to define a strategy for evacuation planning and crisis management. Different types of models exist and may be used depending on the level of accuracy and representation of the flow we search for. But we can identify three main families of flow modelling : microscopic, macroscopic and hybrid models.
In Le Havre Agglomeration (CODAH), we are faced to different types of natural and industrial hazards; 33 establishments are classified SEVESO[2]. In this context, the Major Risk Management Direction team (DIRM) of CODAH has developed a model which estimates the nocturnal and diurnal exposed population allocation PRET-RESSE (Bourcier and Mallet, 2006). So, we have ventilated the day / night population inside buildings. The model was able to locate people during the day both in their workplace and their residence (the unemployed and retirees). Although the model is able to locate the diurnal and nocturnal population, it remains static because it does not take into account the daily movement of people and the road network utilization.
The using of road network by vehicles with diffrent behaviors can generate a danger especially in case of evacuation situations. So, it is important to us to detect congested areas that may be more vulnerable than others. In the literature, many clustering algorithms have been applied on a graph to detect organisations according to a predefined criteria as the connectivity, the distance or capacity of edges between adjacent nodes. Most of these models are static and do not meet the dynamic due to graph evolution.
Since CODAH can be seen as a graph on which we have vehicles flow that evolve during time, we develop a dynamic algorithm which dynamically detects organizations according to the traffic state. The algorithm belongs to collective intelligence algorithms and can be referred to as an ant algorithm. the result of this algorithm must be visualized, as a dynamic map, in the Geographical Information System (GIS) of CODAH to help decision makers in assessing the danger related to road network use by vehicles.

[2]Directive SEVESO is an European directive, it lays down to the states to identify potential dangerous site. It intends to prevent major accidents involving dangerous substances and limit their consequences for man and the environment, with a view to ensuring high levels of protection throughout the Community.

2 CODAH as a complex system

We are intersted in detecting vulnerability related to the road network use in CODAH and we are working with DIRM team to setup a decision making tool assessing this vulnerability. In this context, it is very important to understand the dynamic related to the road network use by vehicles. In fact, a vehicle tends to follow the same path taken by other vehicles to evacuate a dangerous area. In this case, any perturbation (accident...) may cause traffic jams causing serious problem to evacuees. So, we can see CODAH as a system in which the environemnt may influence on evacuation to reduce or not the flow fluidity(existence or not of safe refuge and emergency exits, routes traffic direction) and vice versa (a fire or an accident may cause damages and change the environment).
For us, CODAH can be seen as a complex system in perpetual evolution, in which vehicles interact between them and with environment (vehicles acts on environment which in turn influences vehicles).

Complex system	**CODAH**
A big number of entities interacting with some others (dispersed interactions)	Vehicles on the same route interact but not necessarily with vehicles on other routes
Absence of global control	We have some local rules (traffic lights, priorities on intersections) but no global control
There are cross-hierarchical organizations with many interactions	An organization of rescue service interacts with vehicles in a dangerous area
Continual adaptation, learning and thus evolution	Vehicles learn about congestions in some zones at peak hours in the journey and try to avoid them
Perpetual evolution with appearing and disappearing of organizations	Traffic jams can appear and disappear in time and at different places
The system is far from equilibrium dynamics (no equilibrium or transition from an equilibrium to another one.	Passing from a normal situation to a traffic jam and vice versa

Table 1: CODAH meets complex system properties defined by Arthur *et al.*

A complex system is characterized by numerous entities of the same or different nature that interact in a non-trivial way (non-linear, feedback loop ...); the global emergence of new properties not seen in these entities : Properties or evolution cannot be predicted by simple calculations. According to Arthur, Durlauf and Lane (**?**), a complex system has six properties. Table **??** shows these properties and shows that CODAH can be seen as complex system with a big number of vehicles interacting on the road network.

3 Traffic approaches

Traffic modelling becomes a point of interest for different researchers coming from different fields. It may help decision makers to resolve congestion problems by introducing a set of rules related to environment and different behaviours in entities (vehicles, pedestrians...). In risk context, it helps decision makers to have a better idea about vehicles distribution in space and time to assess vulnerability and to prepare evacuation plans. It can be seen as a complex system in which we have an environement (graph, grid...) and entities in interaction. We can identify three big families of traffic modelling:

- Macroscopic models : The macroscopic model is based on the analogy between vehicular traffic and the fluid flow within a canal.
- Microscopic models : The microscopic model details the behavior of each individual vehicle by representing interactions with other vehicles and in general by using a spatialization. Interactions are generally modeled by car following rules. We can also refine the model by adding passing lanes rules, priority rules on roads intersection, traffic lights, accidents... So, vehicles with different behaviors interact between them and organizations (traffic jams) may appear. With a top down approach, we can examine the global behavior of the system and try to locally modify the environment, when necessary, to ameliorate the system. Recently, some researchers worked on a traffic model in which we have to simulate an urban zone with vehicles and pedestrians. In (**?**), a model is adopted combining cellular automata and Agent Based Model (ABM). For the vehicles, two-way traffic with turning movements are adopted with simples rules of acceleration and deceleration (due to other vehicles, intersections, or pedestrians). For the pedestrians, they have destination points, influenced by marked crosswalk, walk on sidewalk or street, cross if they feel safe (no near vehicles). Accidents may occur if a vehicle hasn't sufficient time to stop when pedestrian crosses the street (signalized crossing or not) and the severity of an accident depends on vehicle speed. It was difficult to compare results with any verified model because, according to the authors, there was no previous study which has been done on such interactions.

- Hybrid models : The hybrid model allows coupling the two types of dynamics flow models within the same simulation. Several works have already borrowed this direction (**???**), however, this approach is relatively new and very few have adopted it to our knowledge (**?**). In risk context, the use of a hybrid model is very important especially when dealing with large volume of data : changing the scale from micro to macro in a region where we haven't a crisis situation (everything is normal) allows to economize the computation and the change from macro to micro in a critical situation allows to zoom and detect the behaviors and interactions between entities in danger.

4 Risk modelling

Traditional methods evaluating the risk for population do not generally treat evacuees behavior (e.g. initial response to an evacuation, travel speed, family interactions / group, and so on.); they describe prescriptive rules as the travel distance. The computerized models offer the potential to evaluate the evacuation of a neighborhood in emergency situations and overcome these limitations (**?**), even than in panic situation, the human behaviour can not be easily predictable and modeled.
In this context, Provitolo (**?**) studied the spread of the panic of a group of individuals in danger situation to non panicked people. For the author, panic generally results from the lack of coordination and dialogue between individuals. She used a dynamic system (differential equations) and STELLA software to simulate the behavior of individuals. Several simulation scenarios were presented and showed that the emergence of panic has not occurred in all scenarios when changing some parameters. The emergence depends on the rate of transmission from a population susceptible to panic to a panicked population , the time taken to return to a normal behavior (the population is more panicked after the disaster) and the number of initially panicked people. The principle of resilience was also discussed and represents the time that the system makes to return to its initial state after a period of instability due to a disaster.
Cova and Church (**?**) opened the way on the study based on geographic information systems to evacuate people. Their study identified the communities that may face transport difficulties during an evacuation. Research has modelled the population by lane occupation during an evacuation emergency using the city of Santa Barbara.
An optimization based model (graph partitioning problem) was realized to find the neighborhood that causes the highest vulnerability around each node in the graph and a vulnerability map around nodes in the city was constructed. A constructive heuristic has been used to calculate the best cluster around each node.

This heuristic was developed in C and the result was displayed on a map (with ArcInfo).
Nevertheless, in this approach, we predefine the maximum number of nodes in a neighborhood (or a community), which may not always be realistic and does not take into account the traffic evolution during the calculation of critical neighborhoods. So, the vulnerable neighborhoods don't evolve according to traffic state.

5 Communities detection

Our aim is to identify communities in graphs according to a predefined criteria and without any fixing of number of communities and the number of nodes per community.

5.1 Some definitions

For Newman et al, a community is a set of nodes strongly linked to each other and more weakly linked to the outside world, according to a predefined criteria, but without any fixing of the nodes number in each community (**?**).
Radicchi (**?**) proposes two possible definitions to quantify Newman definition:

- Community C in a Strong Sense : $d_C^{out}(i) > d_{\bar{C}}^{out}(i), \forall i \in C$.
  A node belongs to a strong community if it has more connections within the community than outside.
- Community C in a Weak Sense : $\sum_{i \in C} d_C^{out}(i) > \sum_{i \in C} d_{\bar{C}}^{out}(i), \forall i \in C$. A community is qualified as weak if sum of all degrees inside is more important than sum of degrees towards the rest of the graph.

$d_C^{out}(i)$ is the exiting edges number from a node i belonging to the community C towards the nodes of the same community.
$d_{\bar{C}}^{out}(i)$ is the exiting edges number from a node i belonging to the community C towards the nodes of other communities. Finding organizations is a new field of research (**??**). Recent works identify two big families in organization detection in large graphs.

- Graph partitioning : a NP-complete problem (**?**)which consists in grouping nodes in fixed number of partitions with a fixed size, while minimizing the number of inter-partitions connections.
- Data partitioning or hierarchical clustering : based on statistical methods to analyse data and regroup similar data by using some similarity distance. Many of works used this type of models (**??**).

Stochastic approaches and collective explorations are generally based on random walk in graphs (**?**). A random walk is a simple stochastic process. It is a formalization of the intuitive idea that we can explore a graph if we walk from an initial node and choose a next adjacent node randomly. To detect organizations, the main idea lays on the fact that small random walks tend to be trapped into them (**??**). A dissimilarity index (**?**) between two nearest neighboring vertices of the graph is measured by a random walk. It is used to determine if nearest-neighboring vertices must be in the same organization. It integrates both the local and the global structural information of the given graph and bias can be introduced, injecting numerical pheromones is an example.

5.2 Proposed algorithm

Our aim is to identify communities in graphs, i.e. dense areas strongly linked to each other and more weakly linked to the outside world. If the concept of communities in a graph is difficult to define formally, it can be seen as a set of nodes whose internal connections density is higher than the outside density without defining formal threshold (**?**). Thus the goal is to find a partition of nodes in communities according to a certain predefined criteria without fixing the number of such communities or the number of nodes in a community.
Interesting works were developed in the literature on the detection of structure in large communities in graphs (**????**).
In our problem, we look for a self-organization in networks with an algorithm close to the detection of communities in large graphs and belonging to collective intelligence algorithms. Organizations connect elements, events or individuals by interrelations so that they become components of a whole. They assume the solidarity and robustness of these links, and ensure that the system will eventually be long lasting despite random perturbations. The organizations, therefore : transform, produce, tie and maintain. Time is present generating the dynamic and we will try to fight against these organizations, in the case of risk management, to avoid bottlenecks which do not facilitate the evacuations.

Thus we will considered the graph at time t, $G(t) = (V(t), E(t))$ where the edges are weighted, this weight being noted $|e|$ for the edge e and represents the needed time to cross this arc, depending on the current load of the traffic and we try to define a colored dynamic graph $G(t) = (E(t), V(t), C(t))$.

The algorithm used can be referred to as an *ant algorithm* (**?**). Our ant algorithm use several colonies of ants, each of a distinct color. Ants travel inside the graph an lay down pheromones, information that can be detected by other ants. Pheromones are also colored. Ants tend to be repulsed by pheromones of other colors. Furthermore, ants tend to favor edges with important weights.

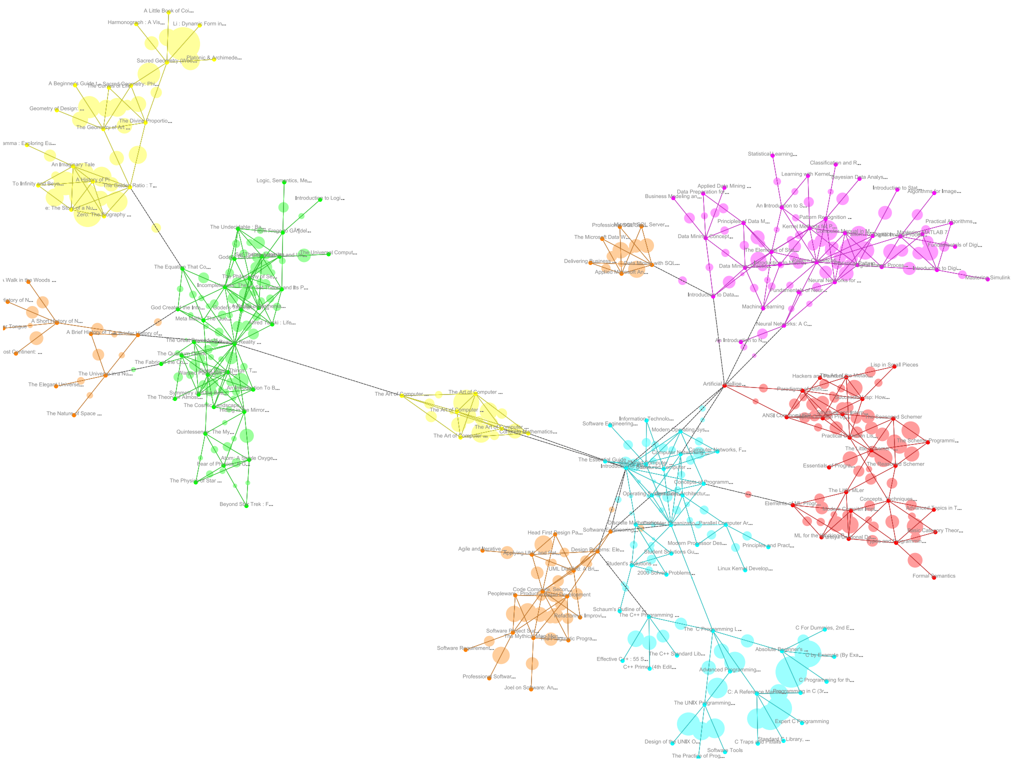

Figure 1: Communities detection in a graph extracted from Amazon website

The colored dynamic graph previously mentionned is defined such that:

- $V(t)$ is the set of vertices at time t. Each vertex v is characterized by:
  - a color $c \in C(t)$,
- $E(t)$ is the set of edges at time t. Each edge e is characterized by:
  - a weight $|e| \in \mathbb{N}^+$ that corresponds to interaction importance between the elements at each end of edge e.
  - a quantity of pheromones of each color.
- $C(t)$ is a set of colors representing the ant colonies at time t.

The algorithm principle is to color the graph using pheromones. Each colony will collaborate to colonize zones, whereas colonies compete to maintain their own colored zone (see figure **??**). Solutions will therefore emerge and be maintained by the ant behavior. The solutions will be the color of each vertex in the graph. Indeed, colored pheromones are deposited by ants on edges. The color of

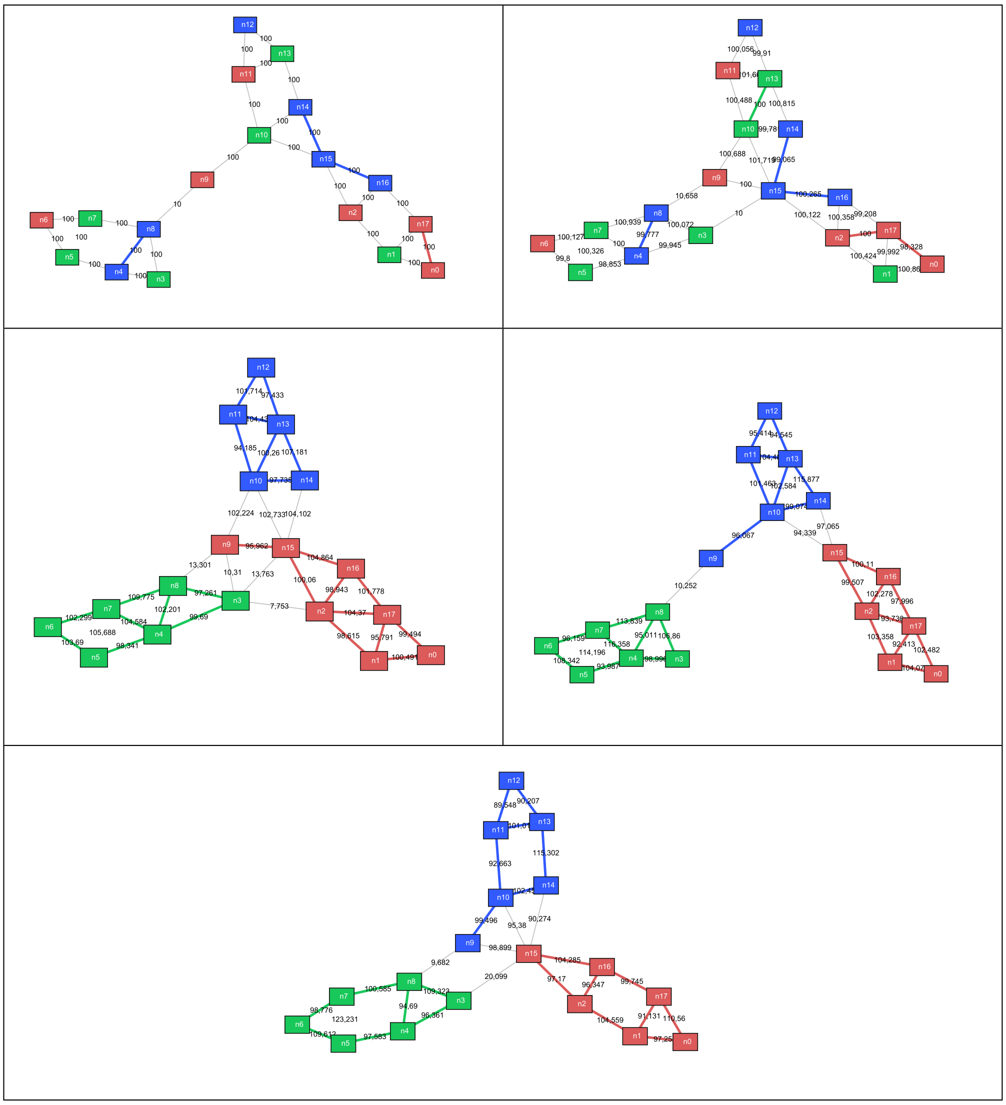

Figure 2: *Example of a dynamic evolution and communities detection.*

a vertex is obtained from the color having the largest proportion of pheromones on all incident edges (see algorithm **??**).

Algorithm 1: Ant behavior

n: current node
t: current time
A: fear of hostile environment threshold
T: resting time
Δt: time counter
if *degree(n)=0* **then**
    Jump randomly on another node
else
    $w \leftarrow$ Sum of all weights on each incident edge to n
    $\tau \leftarrow$ Sum of all pheromones of all colors on each incident edge to n
    $\tau_c \leftarrow$ Sum of pheromones of the ant color on each incident edge to n
    $a \leftarrow \frac{\tau_c}{\tau}$
    if $\Delta t < T$ **then**
        Choose an edge to cross in a weighted random fashion, using edges weight (if available)
        Lay down a small amount of pheromone of the ant color on this edge
        $n \leftarrow$ vertex at the other end of the chosen edge
        $\Delta t \leftarrow \Delta t + 1$
    else
        if $a < A$ **then**
            Jump randomly on another node
            $\Delta t \leftarrow 0$
        else
            Choose an edge to cross in a weighted random fashion, using edges weight (if available)
            Lay down a small amount of pheromone of the ant color on this edge
            $n \leftarrow$ vertex at the other end of the chosen edge

In our problem, we have an interaction between each two local adjacent nodes according to the attraction force that exists between them. At each time step, this force depends in our case on the report of *the actual number of vehicles on the arc between 2 nodes neighbors / vehicles capacity of the arc*. This report was chosen because, in every community, we will have a large number of vehicles which all decide to exit through a single road in the case of a potential danger; this responds well to one of the purposes listed in beginning to have a pessimistic approach in the calculation of vulnerability. The algorithm has the advantage of not allowing the breaking of a link between 2 adjacent nodes to maintain the

structure of the road network. When the traffic evolves, the algorithm detects that and communities can change or disappear as a result of local forces that change between the nodes locally.
In previous work, we applied this algorithm on the Amazon website. Most of the time, when we are on a page that describes some product, several other items are recommended by the site. We see something like *"customers that bought item A also bought B, C, D..."*. The algorithm has detected and separated items in communities depending on their similarity. Currently, we work on the application of the algorithm on a graph with a dynamic flow model (see figure **??**).

6 Dynamic model

Our system consists of two modules as shown in the figure **??**.

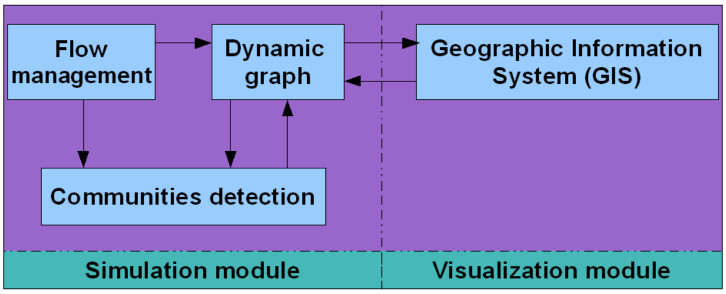

Figure 3: System architecture

6.1 Simulation module

The simulation module contains three components:

- The dynamic graph extracted from the road network layer,
- The flow management component consists of vehicles flow simulator applied on the graph,
- The communities detection component : Its input is the extracted graph and the current flow. It returns the communities that are formed according to the current state of road traffic.

6.2 Visualization module

Recently, some interesting applications have been developed by including the population dynamics, the models of urban growth patterns and land use. In (**?**), the authors emphasize the importance of the collaboration between geographers and multi agents system community to model spatial urban systems. They also underline some conceptual and practical difficulties to such coupling. For computer modelers, this integration provides the ability to have computing entities as agents that are linked to real geographical locations. For GIS users, it provides the ability to model the emergence of phenomena by various interactions of agents in time and space by using a GIS (**?**). So, combining several layers as houses, road network, population... allows us to model different types of agents into a GIS environment. Many researchers have emphasized the need to create vector based Multi Agents System (MAS), which may require the topological data structure provided by using a GIS (**?**). In geography, the representation of a MAS coexists n levels of organizations and use several classes of agents (e.g. Level 1: individuals or companies, Level 2 and three: economic, urban communities).There will be rules at every level and the approach is not necessarily a bottom up one (as in the models of self-organization). . .) (**?**).

In this paper, the visualization module consists of the road network layer integrated into the GIS. This layer represents CODAH network layer. This module communicates with the simulation module: the graph is constructed from this module, which in turn get the simulation result and visualize it as a dynamic vulnerability map.

6.3 Environment modeling

The road network is integrated as a layer in the Geographic Information System (GIS). From this layer, we extract the data by using the open source java GIS toolkit Geotools. This toolkit provides several methods to manipulate geospatial data and implements Open Geospatial Consortium (OGC) specifications, so we can read and write to ESRI shapefile format. Once data road network are extracted, we use the GraphStream tool (**?**) developed within LITIS laboratory of Le Havre to construct a graph corresponding to the GIS network layer. This tool is designed for modeling; processing and visualizing graphs.
The data extracted from network layer contains the roads circulation direction, roads id, roads type, their lengths and geometry.
The extracted multigraph $G(t) = (V(t), E(t))$ represents the road network at time t where $V(t)$ is the set of nodes at t and $E(t)$ the set of arcs at t. We deal with a multigraph because we have sometimes more than one oriented arc in the same direction between two adjacent nodes due to multiple routes between two points in CODAH road network. GraphStream facilitates this task because it is adapted to model and visualize multigraphs. In the constructed multigraph:

- The nodes represent roads intersections,
- The arcs represent the roads taken by vehicles,
- The weight on each arc represents the needed time to cross this arc, depending on the current load of the traffic
- Dynamic aspect relates to the weights of the arcs, which can evolve in time, according to the evolution of the fluidity of circulation.

We have also constructed a Voronoi tessellation (Thiessen polygon) around nodes and projected the population in buildings on these nodes. The population in buildings is extracted from PRET-RESS model.

6.4 Vehicles flow

We have adopted a macroscopic model in which flows circulate normally (without accidents) in order to establish a dynamic pessimistic vulnerability map. In evacuation situations (when an accident occurs), this macroscopic model is not relevant to the reality because most of people panic and take the same road to exit the dangerous area. Hence, it is important to develop a micro approach with multiscale description (from micro to macro and vice versa during the simulation) to simulate scenarios of danger in real time (accidents, behavior of drivers, vehicles interactions…).

7 Conclusion

The proposed method allowed us to dynamically detect organization in large graphs. We will apply this algorithm on the graph extracted from CODAH road network layer. It will serve decision makers to estimate the risk due to the use of the road network by vehicles and categorize CODAH areas by their vulnerability. We will complete our work by using real traffic data retrieved from a displacements survey with CODAH population which will help us to better locate people during the day and therefore having a more realistic vulnerability dynamic map.

References

R Albert and A L Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002.

W.B Arthur, S. Durlauf, and D.A. Lane. *The Economy as an Evolving Complex System II*, chapter Introduction : Process and Emergence in the Economy, pages 1–14. Addison-Wesley, Reading, Mass., 1997.

E. Bourrel and V. Henn. Mixing micro and macro representation of trafic flow : a first theoretical step. In Proceeding of the 9th Meeting of the Euro Work-ing Group on transportation, 2002.

A. Bretagnole, E. Daude, and D. Pumain. From theory to modelling : urban systems as complex systems. *Cybergeo : Revue européenne de géographie*, 335:26p, 2006.

C.J.E Castel. Developing a prototype agent-based pedestrian evacuation model to explore the evacuation of king's cross st pancras underground station. pages 427–437, London, 2006. Center for advanced spatial analysis (university college London): working paper 108, ACM.

R.L. Church and T.J. Cova. Modelling community evacuation vulnerability using gis. *International Journal of Geographical Information Science*, 11(8):763–784, 1997.

A. Clauset, M. E. J. Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical review. E 70*, 6:066111.1–066111.6, 2004.

E. Daudé. *Systèmes multi-agents pour la simulation en géographie : vers une Géographie Artificielle*, chapter 13, pages 355–382. in Y. Guermont (dir.), Modélisation en Géographie : déterminismes et complexités, Hermès, Paris, 2005.

M Dorigo and T Stützle. *Ant Colony Optimization*. MIT Press, 2004.

Antoine Dutot, Frédéric Guinand, Damien Olivier, and Yoann Pigné. Graphstream: A tool for bridging the gap between complex systems and dynamic graphs. In *EPNACS: Emergent Properties in Natural and Artificial Complex Systems*, 2007.

M.R. Garey and D.S. Johnson. *Computers and Intractability : A Guide to the Theory of NP-Completeness*. W H Freeman and Company, New York-San Francisco, 1979.

A. Godara, S. Lassarre, and A. Banos. Simulating pedestrian-vehicle interaction in an urban network using cellular automata and multi-agent models. In Springer, editor, *Traffic and Granular Flow'05*, pages 411–418, 2007.

A. Hennecke, M. Treiber, and D. Helbing. Macroscopic simulation of open systems and micro-macro link. *M. Schreckenberg, D. Helbing, H. J. Herrmann, editor, Traffic and Granular Flow: Social, Traffic, and Granular Dynamics*, 2: 383–388, 2000.

M.S. El Hman, H. Abouaäissa, D. Jolly, and A. Benasser. Simulation hybride de flux de trafic basée sur les systèmes multi-agents. In 6e Conférence Francophone de MOdélisation et SIMulation - MOSIM 06, 2006.

A.K. Jain and R.C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, 1988.

A.K. Jain, M. N Murty, and J. Flynn. Data clustering : a review. *ACM Comput. Surv.*, 31(3):264–323, September 1999.

L. Magne, S. Rabut, and J. F. Gabard. Toward an hybrid macro and micro traffic flow simulation model. In INFORMS spring 2000 meeting, 2000.

R. Najlis and M. J. North. Repast for gis. University of Chicago and Argonne National Laboratory, IL, USA, 2004. In Proceedings of Agent 2004 : Social Dynamics: Interaction, Reflexivity and Emergence.

M. E. J. Newman. Detecting community structure in networks. *Eur. Phys. J.*, 38: 321–330, 2004a.

M. E. J. Newman. Fast algorithm for detecting community structure in networks. *Phys. Rev.*, 69:066133, 2004b.

M. E. J Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys*, 74(3):036104, Sep 2004.

D. C. Parker. Integration of geographic information systems and agent-based models of land use : Challenges and prospects. In *Maguire, D., J. M. F., Goodchild, and M., Batty, (Eds), GIS, Spatial Analysis and Modelling, Redlands, CA: ESRI Press.*, 2004.

P. Pons. Détection de structures de communautés dans les grands réseaux d'interactions. Giens, France, 2005. Septièmes Rencontres Francophones sur les aspects Algorithmiques des Télécommunications.

P. Pons and M. Latapy. Computing communities in large networks using random walks. In *In Proceedings of the 20th International Symposium on Computer and Information Sciences (ISCIS'05) Vol. 3733 of Lecture Notes in Computer Science Istanbul, Turkey*, pages 284–293, October 2005.

P. Pons and M. Latapy. Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, 10(2):191–218, 2006.

D. Provitolo. A proposition for a classification of the catastrophe systems based on complexity criteria. In *European Conference Complex Systems-EPNACS'07, Emergent Properties in Natural and Artificial Complex Systems, Dresden*, pages 93–106, 2007.

F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi. Defining and identifying communities in networks. In *Proceedings of the National Academy of Sciences*, volume 101, pages 2658–2663, 2004.

B. Tadic. Exploring complex graphs by random walks. In *AIP Conference Proceedings*, volume 661, pages 24–27, 2003.

H. Zhou. Distance, dissimilarity index, and network community structure. *Physical Review E*, 67:10, 2003.