DANCer: Dynamic Attributed Network with Community Structure Generator

Oualid Benyahia, Christine Largeron, Baptiste Jeudy, Osmar R Zaïane

DANCer: Dynamic Attributed Network with Community Structure Generator

Oualid Benyahia[1], Christine Largeron ✉[1], Baptiste Jeudy[1], and Osmar R. Zaïane[2]

[1] Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, SAINT-ETIENNE, France.
oualid.benyahia,christine.largeron,baptiste.jeudy@univ-st-etienne.fr
[2] Department of Computer Science, University of Alberta, Edmonton, Canada.
zaiane@cs.ualberta.ca

Abstract. We propose a new generator for dynamic attributed networks with community structure which follow the known properties of real-world networks such as preferential attachment, small world and homophily. After the generation, the different graphs forming the dynamic network as well as its evolution can be displayed in the interface. Several measures are also computed to evaluate the properties verified by each graph. Finally, the generated dynamic network, the parameters and the measures can be saved as a collection of files.

Keywords: Social network mining, Attributed graph, Synthetic data generator

1   Introduction

The proliferation of complex information networks in diverse fields of application has led to the proposal of a panoply of methods to analyze and discover relevant patterns in these networks. However, evaluating these methods and the comparison of the different approaches are not very easy due to the lack of large real networks with ground truth freely accessible to researchers. The alternative consists in using synthetic data provided by generators. There is a large bibliography regarding generation for static graphs, including the classic Erdős-Rényi (ER) model which generates random graphs or the Barabási-Albert (BA) model that generates random scale-free networks, but very few generators allow the construction of evolving graphs, exhibiting or not a community structure and, none of them takes into account the attribute values of the vertices. The interest of community detection, link prediction and more generally pattern discovery in dynamic networks where vertices are associated with attributes led us to develop the generator DANCer for attributed dynamic graphs with embedded community structure. This generator is an extended version of a previous generator, ANC dedicated to static graphs [1].

2 Model

An attributed dynamic network generated by DANCer is represented by (1) a sequence of T attributed graphs $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$, $i \in \{1, \ldots, T\}$, where $\mathcal{V}_i$ is a set of vertices, $\mathcal{E}_i$ a set of undirected edges and where for each vertex $v \in \mathcal{V}_i$ and each real attribute $A \in \mathcal{A}$, v_A denotes the attribute value of A assigned to vertex v and (2) a sequence of T partitions $\mathcal{P}_i$ of $\mathcal{V}_i$, $i \in \{1, \ldots, T\}$ which gives a community for each vertex in the corresponding graph $\mathcal{G}_i$, $i \in \{1, \ldots, T\}$. Each partition allows to define a community structure on a graph (i.e., the network at a single timestamp) in such a way that the nodes are grouped into sets densely connected and relatively homogeneous with regard to the attributes, while they are less connected to vertices belonging to other groups, and less similar with regard to their attributes.

The generation of the network is carried out in two phases. In phase one, an initial graph $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ is built while respecting the well-known network properties such as preferential attachment, small world or homophily and, in the second phase, this initial graph is modified through two kinds of operations. The first set of operations, called micro operations, consist in removing or adding vertices and edges or updating their attributes whereas the second kind of operations is applied on the communities, i.e., at a macro level. They consist in (1) migrating members of a community to either a new community or an existing one, (2) splitting a community into two new sub-communities and (3) merging two existing communities into a single one.

3 Software Overview

The user interface has three panels as shown in Fig. 1. In the parameter panel, the user selects the dynamic generator parameters presented in Table 1[3]. Note that a seed is used for the random number generator. It can be saved to reproduce exactly the same network.

The visualization panel allows to display the generated network and its dynamic evolution. Each graph in the sequence can be selected with a timestamp scrollbar and viewed separately (Fig. 1). This panel can also display the size and the evolution of the different communities in the sequence of graphs according to the macro dynamic operations (split, merge and migrate) (see Fig. 2).

The sequence of attributed graphs is built while preserving properties of real networks and several measures, like modularity, clustering coefficient, diameter, expected and observed homophily or within inertia rate are computed on each graph of the dynamic network to describe its properties[4]. The changes in these different measures on the sequence of graphs are presented at the bottom of the interface in the measure panel (Fig. 3).

---

[3] The reader is referred to [1] for the static network generator parameters `http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0122777`.

[4] see [1] for a more detailed presentation of the properties and corresponding measures.
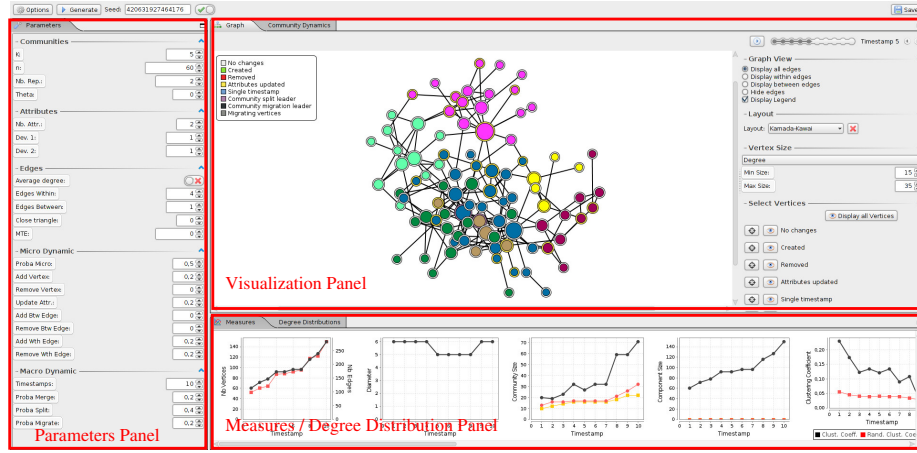
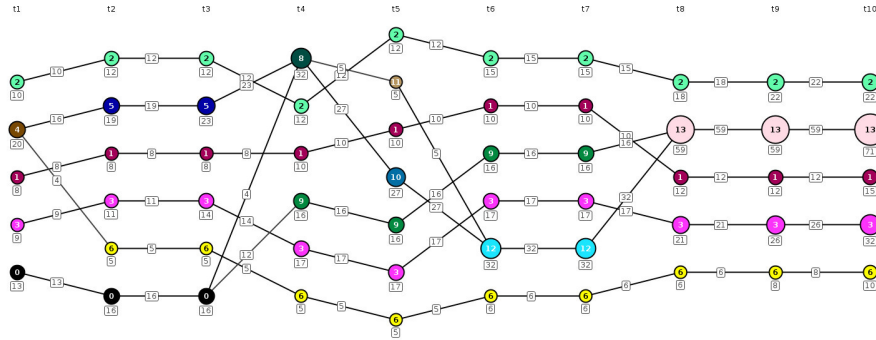**Fig. 1.** User interface of the generator DANCer.

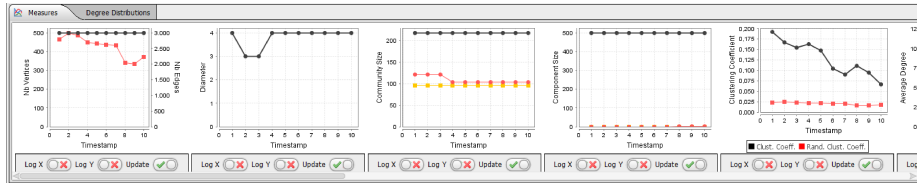**Fig. 2.** Community dynamics display in visualization panel.

Fig. 3. Measures panel of the generator.

The bottom of the user interface includes also a panel displaying the distribution of vertex degrees on each graph of the sequence as shown in Fig. 4.

The generated dynamic network can be saved as a collection of files. For each graph of the sequence, a file indicates the composition of the graph (vertices and edges) and a "**parameters**" file enumerates all the parameters used by the

Parameter	Domain	Description
Micro operations		
Proba Micro	$[0,1]$	A threshold to select if the micro dynamic updates are performed or not
Add Vertex	$[0,1]$	Ratio defining the number of vertices inserted
Remove Vertex	$[0,1]$	Ratio defining the number of vertices removed
Update Attr.	$[0,1]$	Ratio defining the number of attributes updated
Add Btw. Edges	$[0,1]$	Ratio defining the number of between edges inserted
Remove Btw. Edges	$[0,1]$	Ratio defining the number of between edges removed
Add Wth. Edges	$[0,1]$	Ratio defining the number of within edges inserted
Remove Wth. Edges	$[0,1]$	Ratio defining the number of within edges removed
Macro operations		
$P_{removeEdgeSplit}$	$[0,1]$	Proba. to remove an edge between two vertices in the previously same community when splitting a community
Timestamps	$\mathbb{N}^+$	Number of graphs generated
Proba Merge	$[0,1]$	Probability to perform the merge operation
Proba Split	$[0,1]$	Probability to perform the split operation
Proba Migrate	$[0,1]$	Probability to perform the migrate vertices operation

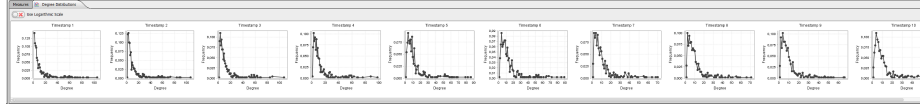**Table 1.** Description of the dynamic network generator parameters

Fig. 4. Degree distribution panel.

generator. The graph measures and community dynamics can also be saved in separated files.

4 Conclusion

The software **DANCer** and a detailed user manual[5] are available under the terms of the GNU Public Licence. Note that our generator can trivially be extended to produce multiplex networks, where all nodes are omnipresent in all levels and intra-level edges connect the representations of a node from one level to the other. This conversion is possible by simply converting each timestamp graph into a layer of the multiplex network and adding the necessary intra-level edges.

References

1. Largeron C., Mougel P.N., Rabbany R., Zaïane O.R. Generating Attributed Networks with Communities. PLoS ONE 10(4): e0122777 (2015)

---

[5] http://perso.univ-st-etienne.fr/largeron/DANCer_Generator/