Point of View Based Clustering of Socio-Semantic Networks

Juan David Cruz*, Cécile Bothorel*, François Poulet**

*Département LUSSI – Télécom – Bretagne
{juan.cruzgomez, cecile.bothorel}@telecom-bretagne.eu,
http://www.telecom-bretagne.eu/
**Université de Rennes 1 – IRISA
francois.poulet@irisa.fr
http://www.irisa.fr/texmex/

1 Introduction

Socio–semantic networks contain two different types of information, the social relationships and information related to the actors such as thematic profiles.

In order to analyze such augmented network from different perspectives, we propose to influence the communities detection process with semantic information, the clustering process is divided into two phases. In the first one, the point of view is clustered using Self – Organizing Maps (SOM) (Kohonen (1997)) to obtain groups based on the similarity of the node features and changing the weights of the graph according to the semantic groups found. Then, in the second phase, a classic community detection algorithm is used, based on the topology of the network.

2 Using the Point of View to Influence the Clustering

A social network can be represented as a non–directed graph $G(V, E)$ where V is the non-empty set of vertices representing actors and E is the set of edges representing the relationships among them.

Given a graph $G(V, E)$, $\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$ is a partition of the set V into k non–empty disjoint subsets C_i.

Let F_V be the set of semantic features of the actors of the social network, which can be represented by a matrix of size $|V| \times |F_V|$.

Let $F_V^* \in \mathcal{P}(F_V) \setminus F_V$, where $\mathcal{P}(A)$ is the powerset of the set A, be a non–empty set of features to be used to define the point of view PoV.

For each vertex $v_i \in V$ we assign a binary vector $\xi_i = v_i \times F_V^*$ of size $|F_V^*| = f$. If the vertex i has the feature $p, 1 \leq p \leq f$ from F_V^*, then $\xi_{i,p} = 1$ or 0, otherwise.

A point of view is defined as the set of all instances derived from the set F_V :

$$PoV_{F_V^*} = \bigcup_{i=1}^{|V|} \xi_i \tag{1}$$

1. **Phase 1 : Semantic Clustering** Using SOM we group the nodes according to the similarities of their features. For training the SOM network, each instance ξ_i is used as an input pattern.

   The outcome of the SOM algorithm is a partition $\mathcal{C}_{SOM}$ of the nodes assigned to the neurons.

   For each pair of neighbor vertices $v_i, v_j, \forall i \neq j \in V$, the weight of the edge $e(v_i, v_j)$ is changed according to the Euclidean distance of the PoV instances corresponding to each node by :

$$w_{ij} = 1 + \alpha \left(1 - d\left(\mathcal{N}_{ij}\right)\right) \delta_{ij} \tag{2}$$

   where $\alpha \geq 1$ is a constant value, $d\left(\mathcal{N}_{ij}\right)$ is the distance between the neurons i and j, and $\delta_{ij} = 1$ if v_i and v_j belong to the same partition in $\mathcal{C}_{SOM}$, $\delta_{ij} = 0$ otherwise.

2. **Phase 2 : Structural Clustering and Community Detection** This algorithm uses the modularity Q, which is proposed by Newman et Girvan (2004) as quality measure.

   After the weights are changed according to Equation 2, a partition $\mathcal{C}_{SOM \rightarrow FU}$ is computed using the Fast Unfolding algorithm proposed by Blondel et al. (2008), a classic topology based algorithm to find communities in graphs.

   This new partition $\mathcal{C}_{SOM \rightarrow FU}$ contains the final set of communities, which has both the semantic information and the structural information.

3 Conclusion and Future Work

The information contained into a socio-semantic network, such as Twitter, is tied to certain features of the actors.

The proposed method allows to analyze Twitter information from different points of view like the olympic games communities on geolocated comments.

Assigning weights derived from the results of the semantic clustering to the edges, the semantic information is included into the community detection process and the two types of data are merged to find and visualize a social network from a selected point of view.

For future work we will also continue the study of the influence of the point of view in the community detection process including the definition of points of view from the graph's edges. Additionally, we plan to work on the development of a visualization algorithm for hierarchical social networks.

Références

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, et E. Lefebvre (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics : Theory and Experiment 2008*(10), P10008 (12pp).

Kohonen, T. (1997). *Self-Organizing Maps*. Springer.

Newman, M. E. J. et M. Girvan (2004). Finding and evaluating community structure in networks. *Physical Review. E, Statistical Nonliner and Soft Matter Physics 69*(2), 026113.