Network node label acquisition and tracking *

Sarvenaz Choobdar, Fernando Silva, and Pedro Ribeiro

CRACS & INESC-Porto LA
Faculty of Science, University of Porto, Portugal
{sarvenaz,fds,pribeiro}@dcc.fc.up.pt

Abstract. Complex networks are ubiquitous in real-world and represent a multitude of natural and artificial systems. Some of these networks are inherently dynamic and their structure changes over time, but only recently has the research community been trying to better characterize them. In this paper we propose a novel general methodology to characterize time evolving networks, analyzing the dynamics of their structure by labeling the nodes and tracking how these labels evolve. Node labeling is formulated as a clustering task that assigns a classification to each node according to its local properties. Association rule mining is then applied to sequences of nodes' labels to extract useful rules that best describe changes in the network. We evaluate our method using two different networks, a real-world network of the world annual trades and a synthetic scale-free network, in order to uncover evolution patterns. The results show that our approach is valid and gives insights into the dynamics of the network. As an example, the derived rules for the scale-free network capture the properties of preferential node attachment.

Keywords: Network Characterization, Node labeling, Clustering, Association Rules

1 Introduction

Advances in information technology led the world activity to become very much centered on information data. The explosive growth in data that we are witnessing naturally opens an enormous opportunity for researchers to develop new methodologies to dynamically extract useful information and knowledge from the data. Real life data inherently contains structural information on objects and their relationships. This structure can be modeled with networks, or graphs, that are abstract representations of a set of nodes and the connections between them.

Most real world networks are complex, in the sense that they present non trivial topological features. Research on complex network data analysis has been very prolific and a large variety of characterization methodologies emerged, such as node classification [23], graph clustering [13, 25], frequent subgraph mining [32, 16] or network motifs discovery [24]. These approaches treat the network as a

* In Proceedings of EPIA 2011, Springer, LNAI 7026, 418–430, 2011.

static object. However, many networks are intrinsically dynamic and change over time. Only recently has the research community started to analyze the temporal evolution of networks [22, 4, 21, 7, 6]. Most of these studies have characterized network structure by directly examining the topology of the network. Nevertheless, more indirect methods that use network measurements such as degree centrality or clustering coefficient can be a rich source of information [9]. Our aim is precisely to use the evolution of these kind of metrics to study the dynamics of networks that change over time.

We propose a novel two-phase general methodology designed to characterize time evolving networks. First we group and classify nodes based on their role in the network, and then we track changes and find patterns on how this classification evolves.

The first phase involves looking at the network from a static point of view and creating a node classification. We can either use a predefined label for each node, or an automated classification of nodes based on their local properties, such as degree or betweenness, which have been shown to be very fruitful in node characterization [9]. It is however not an easy task to choose a set of these kind of metrics that best generally describes and distinguishes nodes. Costa et al. [10] presented a node label acquisition methodology based on these metrics, but they only identify outliers, that is, singular nodes that are the most different from the others. Instead, we apply these measurements to all nodes in the networks and then employ clustering techniques to group nodes and attribute them labels.

The second phase involves tracking the evolution of node labels over time. We mine association rules using the *apriori* algorithm [2], characterizing the dynamics of the networks in order to uncover emerging patterns that show the appearance, change and disappearance of groups of nodes.

We apply our methodology on a real and on a synthetic complex network in order to demonstrate the validity and usefulness of our approach. We show that our method can discover interesting insights in dynamic networks.

In the remainder of the paper, we start by presenting recent and state of the art work regarding network evolution. After, we describe the proposed methodology in section 3, with all intermediate steps and techniques used. Then, we evaluate our approach in section 4, by applying it to two network datasets. Finally, we draw some conclusions.

2 Related Work

2.1 Graph clustering

Graph clustering is the task of grouping the nodes of the graph into clusters taking into consideration the edge structure of the graph in such a way that there should be many edges within each cluster and relatively few between the clusters [25].

2.2 Label acquisition

Label acquisition, as most commonly defined in the literature, involves determining the label for a node in a network that is partially labeled. Normally, it is assumed that at least some of the nodes have a predefined label and only the labels for remaining nodes are predicted using relational classifiers [30].

With networked data, the label of a node may influence the label of a related node. Furthermore, nodes not directly connected may be related through chains of links. This complex dependencies thus suggest that it may be beneficial to predict the label of all nodes simultaneously. Regarding the values of an attribute or attributes for multiple connected nodes for which some attribute values are unknown, a simultaneous statistical assessment is required and this can be done by using collective inferencing [17]. Networked data allow collective inference, meaning that various interrelated values can be derived simultaneously [23]. Macskassy et al. [12] used another source of information in networks that is independent of the available node labels and improved the accuracy of node's label by adding label independent features which include nodes local measurements like degree and betweenness.

All the mentioned studies aim to find the label of a node in a partial labeled network and rely mainly on the available information about label of some of the nodes, they predict labels of nodes instead of assigning the labels. Our work in label acquisition follows the work by Costa et. al. [10], but differs in that we consider all the nodes in the network instead of just the singular node-motifs.

2.3 Node evolution

Different approaches for explaining network evolution have been reported in the literature. Some have focused on the global evolution of networks by an exploratory point of view. Leskovec et al. [22] discovered the shrinking diameter phenomena on time-evolving networks. Backstrom et al. [4] studied the evolution of communities in social networks. Still from an exploratory perspective, Leskovec et al. [21] studied the evolution of networks but at a more local level. Using a methodology based on the maximum-likelihood principle, they investigate a wide variety of network formation strategies, and show that edge locality plays a critical role in evolution of networks[6].

Braha and Bar-Yam [7] described node's centrality changes over time and showed that hubs do not remain a hub for the all time. They use nodes degree over time to computes correlation between pairs of daily networks.

Other recent papers, present algorithmic tools for the analysis of evolving networks. Tantipathananandh et al. [29] focus on assessing the community affiliation of users and how this changes over time. Sun et al. [27], apply the minimum description length (MDL) principle to the discovery of communities in dynamic networks, developing a parameter-free framework. This is the main difference to previous work such as [1, 28]. However, as in [31], the focus lies on identifying approximate clusters of users and their temporal change. No exact patterns are

found, nor is time part of the results obtained with these approaches. Ferlez et al. [11] use the MDL principle for monitoring the evolution of a network.

Network motifs as small subgraphs that show the topological properties of the network have also been used in [19] to monitor temporal changes in the structure of an email communication network. They considered z-score of motif as its significance in the network and trace it over time. The dynamics of network is studied in [7] by calculating the network motifs frequency over the time.

3 Methodology

We address the characterization of network dynamics by tracking the evolution of groups of nodes over time, and deriving rules that explain that evolution. We do this in two phases: one deals with node label acquisition, and the other deals with node evolution pattern discovery. The first phase has multiple steps, as illustrated in figure 1, that include measuring local properties of nodes, determining the proper number of clusters in the network, clustering the nodes, coordinating the clusters of nodes in the network time span and assigning labels to groups of nodes. The second phase comprises defining a time granularity for network pattern detection and mining association rules.

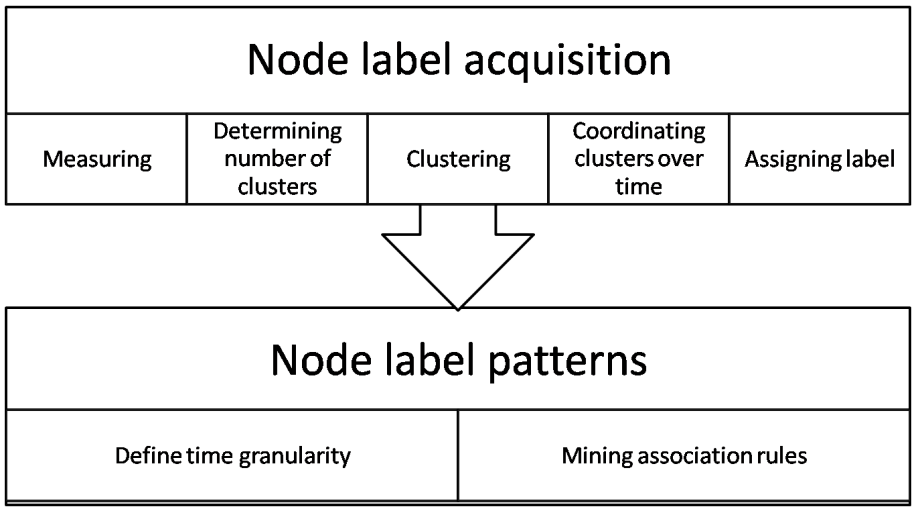

Fig. 1: Proposed methodology

3.1 Node label acquisition

The label of a node is determined based on its properties in the network. The same label is assigned to the nodes that have similar properties. Therefore, first

we select a set of local measurements that best characterize nodes in the network structure, and then we determine the groups of nodes that have similar properties. The same metrics as introduced by Costa et al. [10] are employed as "feature vector" for nodes. This feature vector measures the connectivity of a node in the neighborhood structure. It includes:

- the normalized average degree (r),
- the coefficient variation of the degrees of the immediate neighbors of a node (cv),
- the clustering coefficient (cc),
- the locality index (loc), which is an extension of the matching index and takes into account all the immediate neighbors of each node, instead of individual edges, and
- the normalized node degree (K).

Label assignment requires the grouping of identical nodes, that is, nodes that share the same feature vector. We use multivariate statistics and pattern recognition [18] techniques to find these identical nodes. Clustering is a method widely used for finding groups of objects, called clusters, in the dataset such that the objects in the same group are more similar to each other than they are to objects of other groups. We use the well known k-means clustering algorithm [15], which bases its operation on the euclidean distance between nodes. If two nodes have similar feature vectors, they are clustered into the same group. After clustering the nodes, coherent groups of nodes are derived and, therefore, can be said to have the same role or label in the network.

The number of potential groups of nodes in the network is equal to the number of clusters in the dataset. Determining the actual number of groups in a dataset is a fundamental and largely unsolved problem in cluster analysis. We employ the method by Sugar and James [26], since it does not require parametric assumptions, is independent of the method of clustering, and was shown to achieve excellent results. This method uses a theoretic information approach that considers the transformed distortion curve $d_K^{-p/2}$ [26]. *"Distortion"* is a measure of within cluster dispersion which is a kind of average Mahalanobis distance between the data and the set of cluster centers as a function of the number of clusters, K. This method is called the *"jump method"*. First, it runs the k-means algorithm for different numbers of clusters, K, and calculates the corresponding distortions, $\hat{d}_K$. Then it transforms the distortion by power transformation of $Y = p/2$, where p is the number of dimensions in the dataset. The "jumps" in the transformed distortion are calculated by $J_k = \hat{d}_K^{-y} - \hat{d}_{K-1}^{-y}$. Finally, the appropriate number of clusters for the data is equal to $K^* = argmax_k J_k$.

We use the jump method to determine the groups of nodes in the network at each time instance. The nodes that are members of the same cluster hold the label of that cluster. The labels can be the numbers of the cluster or be defined manually by the domain expert, based on the properties of the cluster's center.

After labeling the nodes in the network over time, we are left with a series of networks whose nodes are independently labeled. For example, it may be the case

that group number one at time t includes nodes with low degree, low clustering coefficient and high neighborhood degree, but at time $t + 1$ it may be that it is group number two that includes nodes with this feature vector. That is, the same feature vector might appear in different groups of nodes at different times. Therefore, a coordination is required for the labels. For the coordination, we use the centers of the detected clusters at each time. We assign the same label to the nodes belonging to two clusters whose centers are close to each other. By considering the Mahalanobis distance of the centers, we can derive an universal label for the nodes of the network for all the time instances.

At the end of this phase, coherent groups of nodes at each time instance are derived and labeled. Therefore, a sequence of labels is generated for each node over time. In the next phase, we attempt to extract from these sequences patterns that explain nodes evolution.

3.2 Node evolution patterns

Having a time-evolving network with labeled nodes is a requirement for determining evolution patterns. Our goal is to find rules that explain transitions of nodes between groups over time. We consider the labels of nodes as items, and thus an itemset in our case is the sequence of node's labels over the time. Frequent itemsets are the sequences or subsequences which have minimum support. Therefore, the patterns of node evolution are the extracted frequent itemsets and association rules.

The Apriori algorithm is a powerful tool for mining associations, correlations, causality and sequential patterns [2, 3, 8, 20]. Association rules mining has two main steps [2]:

1. Finding all sets of items (itemsets) whose transaction support is above a minimum support threshold. The support for an itemset is the number of transactions that contain the itemset. Itemsets with minimum support are called large itemsets, and all others small itemsets.
2. Use the large itemsets to generate the desired rules for every large itemset l and all non-empty subsets of l. For every such subset l_a, output a rule of the form $l_a \Rightarrow (l - l_a)$ if the ratio of support (l) to support (l_a) is at least minimum confidence. We need to consider all subsets of l to generate rules with multiple consequents.

The apriori algorithm generates the frequent itemsets with different time granularity. Patterns of evolutions are generated using a sliding window method that enable us to detect changes at different stages of network lifetime. At each time window, rules with different time granularity are extracted.

4 Experiments

The proposed methodology for automatically assigning labels to nodes and track their evolution over time was implemented in R. We evaluated it on two different

datasets, a network of the world countries' global trade (GDP data) [14], and a synthetic scale free network. We start by describing these datasets in some detail and then present our evaluation results.

4.1 Datasets

The first data set is created from the publicly available Expanded Trade and GDP Data [14]. The data represents the yearly imports and exports, total trade and gross domestic product (GDP) of 196 countries spanning the 52 years from 1948 till 2000. The time series for each country is the proportion of its share in the global economy according to its GDP for that year. The time series for GDP-Norm is the normalized value of each individual annual GDP, divided by the total GDP for all countries during that year. The topology for the graph was created by comparing the yearly total trade for each country and its trade with each of the other countries. If the trade between country A and country B in any given year accounts for more than 10% of either country's total trade for that year, an edge is created between the two countries.

The second dataset is a synthetic scale-free network generated based on the Barabasi-Albert model for graph generation [5]. It is a model of network growth that is based on two basic parameters: growth and preferential attachment. The basic idea is that in the network nodes with high degrees acquire new edges at higher rates than low-degree nodes. An undirected graph is constructed as follows. Starting with m_0 isolated nodes, at each time step $t = 1, 2, \ldots, N$ a new node j with $m \leq m_0$ links is added to the network. The probability that a link will connect j to an existing node i is linearly proportional to the actual degree of node i given by

$$P(k_i) = k_i / \sum_j k_j \tag{1}$$

4.2 Results

Table 1 provides details on the networks used in our experiments, namely the number of nodes in each network, the number of time instances of network evolution, and the number of labels produced with the application of the first step of our method for node label acquisition.

Dataset	# nodes	# times	# labels
GDP	171	52	4
Scale-Free	200	100	7

Table 1: Statistics of used networks

Figures 2 and 3 show the profile of the groups found in each network. The profile depicts the values of the feature vector of each group in the network.

As explained earlier, the feature vector includes the metrics normalized average degree (`r`), coefficient variation of the degrees of immediate neighbors (`cv`), the clustering coefficient (`cc`), the locality index (`loc`), and the normalized node degree (`K`).

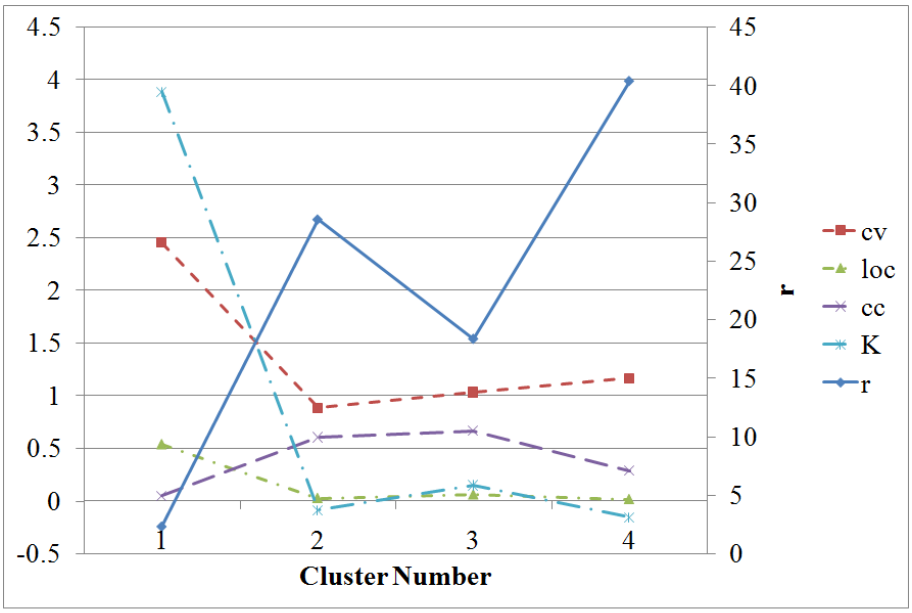

Fig. 2: Profile of feature vector in groups of nodes in the GDP network. The feature of `r` is depicted in the right side of the vertical axis.

GDP network: in this global trade network of countries, our method found four distinguishable groups of nodes. Each group has a different feature vector as illustrated in Figure 2. The first group includes nodes that represent countries with very high degree and many low degree nodes connected to them. Neighbors of these nodes have low degree since the normalized average degree of the immediate neighbors of a node for this group is very low. This means that nodes of group one behave as hubs in the network, that is, as hub countries in global trade, with commercial transactions with many other countries that have a high variation of degree in neighborhood (`cv`). According to the value of `loc` and `cc`, respectively, the locality index and the clustering coefficient, nodes of this group are highly connected in their neighborhood. United States of America, Canada and France are members of this group.

Figure 4 depicts the evolution on the frequency of each label over time in each network. At the initial stages of network evolution, groups with label 1 and 3 are rather common, but they become rare as the network evolves. These

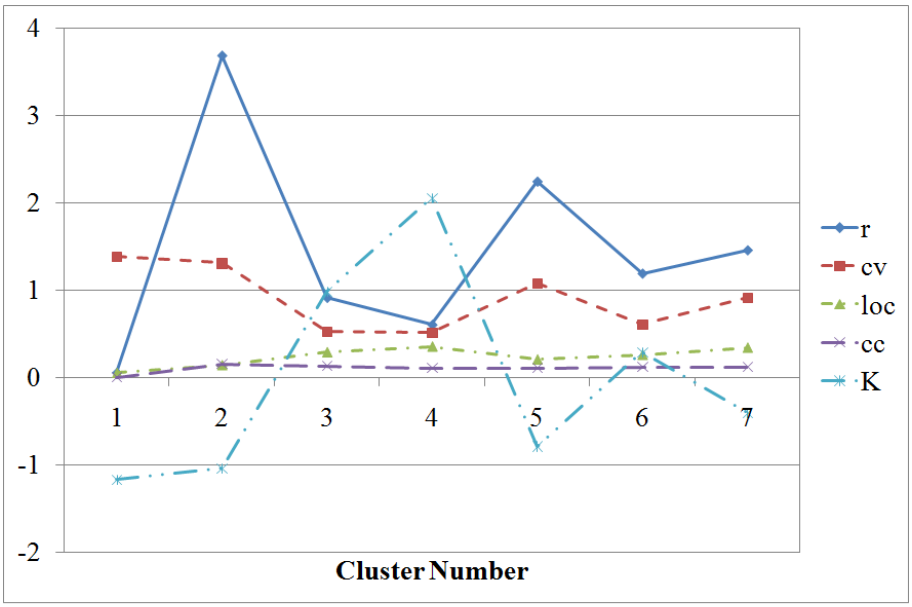

Fig. 3: Profile of feature vector in groups of nodes in the Scale-Free network.

groups have different sizes (number of nodes) at each time step, but they never vanish. Over time, one can notice a transition from group 3 to group 2. After the initial stages of network lifetime, a new group emerges, in this case group number four.

The changes over time in the role of the nodes of certain groups is described by the extracted patterns of node's evolution. Table 2 shows the stronger rules, in terms of support and level of confidence, that are extracted from the label sequence of nodes in GDP network. We used a sliding window to find out the changes in the network. The sliding window (SW) parameter helps to narrow down the search interval to find more precisely rules that describe changes of node's label. The size of the SW can be determined by the Fig. 4 which shows the trend of node's membership. With different SW size, several rules could be found. The significant ones that characterize the appearance and disappearance of the groups are listed in table 2.

For example group four does not exist in the network before time step 18. This pattern of change is detected and described by rule of $\{t14 = 2, t17 = 2\} \Rightarrow \{t18 = 4\}$. This rule says that nodes that were in the second group at time 14 and 17 are likely to change to the fourth group at time 18. The support for this rule is 11%, but its confidence is 87%.

Scale-Free network: this network was generated with 200 nodes and we sampled 100 networks from its evolution time. Our method detected seven different groups

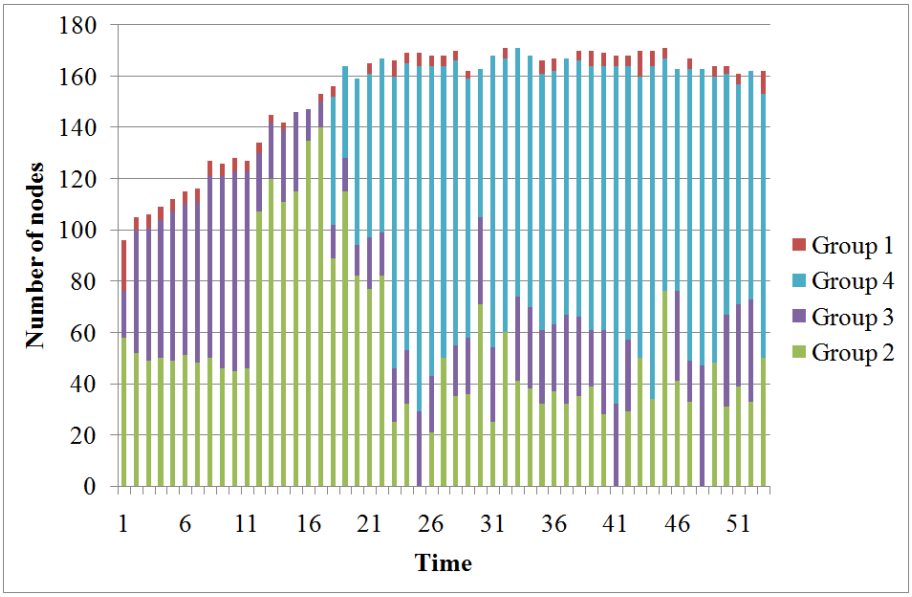

Fig. 4: Frequency of groups in the GDP network over time.

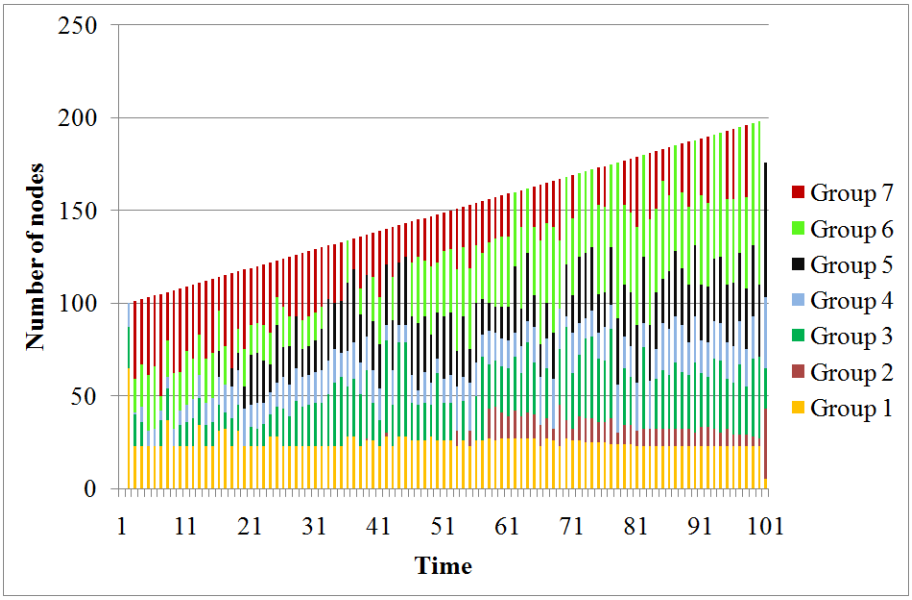

Fig. 5: Frequency of groups in the Scale-Free network over time.

Network	Rule	Support	Confidence
GDP	$\{t1 = 2, t4 = 2, t8 = 3\} \Rightarrow \{t9 = 2\}$	18%	95%
	$\{t14 = 2, t17 = 2\} \Rightarrow \{t18 = 4\}$	11%	87%
	$\{t28 = 4, t29 = 1\} \Rightarrow \{t30 = 3\}$	16%	70%
	$\{t40 = 2, t41 = 3, t42 = 3\} \Rightarrow \{t43 = 2\}$	17%	75%
Scale Free	$\{t16 = 7, t = 18 = 7\} \Rightarrow \{t20 = 5\}$	6%	72%
	$\{t57 = 5\} \Rightarrow \{t60 = 2\}$	11%	82%
	$\{t75 = 6, t76 = 6\} \Rightarrow \{t80 = 3\}$	7%	81%

Table 2: Derived rules for the networks

of nodes with distinct feature vectors, as illustrated in Figure 3. The first group of nodes in this network are those that are weakly connected such that all of their local connectivity properties in the feature vector of this group have the lowest values between the nodes but have a very high variation of degree in their neighborhood (`cv`). A reason for this is that the neighbors of these nodes are mostly low degree nodes that, however, are connected to a hub in the network with very high degree. As shown in Figure 3, these groups have different numbers in each time but never cut down in the network lifetime. Second group are nodes with low degree. This group was formed almost at the middle of network evolution time span (low `K`). They are connected to very high degree nodes (high `c` and high `cv`). This group of nodes appears after the 50th time instance. The third and forth group of nodes are highly connected nodes (high `K` and `cc`) with neighborhood of low degree nodes (low `r` and `cv`). The other three groups of nodes, 5, 6 and 7, are low degree nodes, but the nodes in fifth group are also connected to a hub, which does not happen with the others groups. The sixth group emerges at the beginning stage of the network development and becomes more frequent as time goes on.

Extracted rules in table 2 describe the stronger trends in nodes' transitions between groups. For example $\{t57 = 5\} \Rightarrow \{t60 = 2\}$ shows that nodes in fifth group, after a while, change to the second group. This rule also shows that as time goes on, regarding the generation model of the scale-free network, although the neighborhood of the nodes get more crowded (`r` and `cv` increases), their degree remains low. Nodes in the fifth group have low degree, thus they can not absorb new connections and their degree does not increase.

5 Conclusions

Many networks are intrinsically dynamic and evolve over time. Discovering topological features in these networks is far from an easy task. In this work, we proposed a network characterization method that considers both, a static and a dynamic point of view. It is a two phase methodology that automatically assigns labels to nodes of the network based on their local properties and extracts patterns of nodes evolution. The static view provides a general description of the network through label assignment to groups of nodes. Each group in the net-

work is well characterized by the corresponding feature vector profiling. From a dynamic point of view, the methodology discovers patterns of network evolution at node level. Extracted patterns are general rules that describe how one node's label changes from time to time.

We applied our method to two networks to demonstrate and assess its capabilities. It successfully clusters nodes in groups performing a similar role in the network, labels the groups and, through association rule mining, derives rules that explain, with high confidence, patterns of network evolution. The rules show node transitions between groups as time evolves.

Future research will be pursued to extend this methodology, so that we do not just look at individual nodes but subgraphs in the network. In particular, given our prior work on efficient methods for motifs discovery, we are specially concerned with using subgraph motifs as a metric for network characterization, and then studying network evolution based on such larger entities.

6 Acknowledgments

Sarvenaz Choobdar is funded by an FCT Research Grant (SFRH/BD/72697/2010). The authors are also thankful to Maria Eduarda Silva for the fruitful discussions and suggestions that she made.

References

1. C.C. Aggarwal and P.S. Yu. Online analysis of community evolution in data streams. In *Proceedings of the 15th SIAM International Conference on Data Mining*, page 56. Society for Industrial Mathematics, 2005.
2. R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proceedings of 20th International Conference on Very Large Data Bases*, volume 1215, pages 487–499, 1994.
3. R. Agrawal and R. Srikant. Mining sequential patterns. In *11th International Conference on Data Engineering*, page 3, 1995.
4. L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 44–54, 2006.
5. A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509, 1999.
6. Michele Berlingerio, Francesco Bonchi, Björn Bringmann, and Aristides Gionis. Mining graph evolution rules. In *Machine Learning and Knowledge Discovery in Databases*, pages 115–130. Springer, September 2009.
7. D. Braha and Yaneer Bar-Yam. Time-dependent complex networks: Dynamic centrality, dynamic motifs, and cycles of social interactions. In *Adaptive Networks*, volume 51, pages 39–50. 2009.
8. S. Brin, R. Motwani, and C. Silverstein. Beyond market baskets: Generalizing association rules to correlations. In *ACM SIGMOD Record*, volume 26, pages 265–276, 1997.

9. L. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. Characterization of complex networks: A survey of measurements. *Advances In Physics*, 56:167, 2007.
10. L.F. Costa, F.A. Rodrigues, C.C. Hilgetag, and M. Kaiser. Beyond the average: detecting global singular nodes from local features in complex networks. *EPL (Europhysics Letters)*, 87:18008, 2009.
11. J. Ferlez, C. Faloutsos, J. Leskovec, D. Mladenic, and M. Grobelnik. Monitoring network evolution using MDL. In *IEEE 24th International Conference on Data Engineering*, pages 1328–1330, 2008.
12. Brian Gallagher and Tina Eliassi-Rad. Leveraging label-independent features for classification in sparsely labeled networks: an empirical study. In *Proceedings of the 2nd international conference on Advances in social network mining and analysis*, SNAKDD'08, pages 1–19, Berlin, Heidelberg, 2010. Springer-Verlag.
13. M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *PNAS*, 99(12):7821–7826, June 2002.
14. K.S. Gleditsch. Expanded trade and GDP data. *Journal of Conflict Resolution*, 46(5):712, 2002.
15. JA Hartigan and MA Wong. A k-means clustering algorithm. *Journal of the Royal Statistical Society C*, 28(1):100–108, 1979.
16. Jun Huan, Wei Wang, and Jan Prins. Efficient mining of frequent subgraphs in the presence of isomorphism. In *Proceedings of the Third IEEE International Conference on Data Mining*, ICDM '03, pages 549–, 2003.
17. D. Jensen, J. Neville, and B. Gallagher. Why collective inference improves relational classification. In *Proceedings of the 10th KDD*, 2004.
18. R.A. Johnson and D.W. Wichern. *Applied multivariate statistical analysis*, volume 5. Prentice Hall Upper Saddle River, NJ, 2002.
19. K. Juszczyszyn, P. Kazienko, K. Musial, and B. Gabrys. Temporal Changes in Connection Patterns of an Email-Based Social Network. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, volume 3, pages 9–12, 2008.
20. M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A.I. Verkamo. Finding interesting rules from large sets of discovered association rules. In *Proceedings of the 3rd international conference on Information and knowledge management*, pages 401–407, 1994.
21. J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins. Microscopic evolution of social networks. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 462–470, 2008.
22. J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the 11th ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 177–187, 2005.
23. S.A. Macskassy and F. Provost. Classification in networked data: A toolkit and a univariate case study. *The Journal of Machine Learning Research*, 8:935–983, 2007.
24. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network Motifs: Simple Building Blocks of Complex Networks. *Science*, 298(5594):824–827, October 2002.
25. Satu Elisa Schaeffer. Graph clustering. *Computer Science Review*, 1(1):27–64, 2007.
26. C.A. Sugar and G.M. James. Finding the number of clusters in a dataset. *Journal of the American Statistical Association*, 98(463):750–763, 2003.

27. Jimeng Sun, Christos Faloutsos, Spiros Papadimitriou, and Philip S. Yu. Graphscope: parameter-free mining of large time-evolving graphs. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 687–696, 2007.
28. Jimeng Sun, Dacheng Tao, and Christos Faloutsos. Beyond streams and graphs: dynamic tensor analysis. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 374–383, 2006.
29. Chayant Tantipathananandh, Tanya Berger-Wolf, and David Kempe. A framework for community identification in dynamic social networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 717–726, 2007.
30. B. Taskar, P. Abbeel, and D. Koller. Discriminative probabilistic models for relational data. In *18th Conference on Uncertainty in Artificial Intelligence*, pages 895–902, 2002.
31. N. Vanetik, S.E. Shimony, and E. Gudes. Support measures for graph data. *Data Mining and Knowledge Discovery*, 13(2):243–260, 2006.
32. Xifeng Yan and Jiawei Han. gspan: Graph-based substructure pattern mining. In *Proceedings of the 2002 IEEE International Conference on Data Mining*, ICDM '02, pages 721–, 2002.