As Time Goes by: Discovering Eras in Evolving Social Networks

Michele Berlingerio[1], Michele Coscia[1,2], Fosca Giannotti[1,3], Anna Monreale[1,2], and Dino Pedreschi[2,3]

[1] ISTI - CNR, Area della Ricerca di Pisa, Italy
{name.surname}@isti.cnr.it
[2] Computer Science Dep., University of Pisa, Italy
{coscia,annam,pedre}@di.unipi.it
[3] Center for Complex Networks Research - Northeastern University, Boston, MA

Abstract. Within the large body of research in complex network analysis, an important topic is the temporal evolution of networks. Existing approaches aim at analyzing the evolution on the global and the local scale, extracting properties of either the entire network or local patterns. In this paper, we focus instead on detecting clusters of temporal snapshots of a network, to be interpreted as *eras* of evolution. To this aim, we introduce a novel hierarchical clustering methodology, based on a dissimilarity measure (derived from the Jaccard coefficient) between two temporal snapshots of the network. We devise a framework to discover and browse the eras, either in top-down or a bottom-up fashion, supporting the exploration of the evolution at any level of temporal resolution. We show how our approach applies to real networks, by detecting eras in an evolving co-authorship graph extracted from a bibliographic dataset; we illustrate how the discovered temporal clustering highlights the crucial moments when the network had profound changes in its structure. Our approach is finally boosted by introducing a meaningful labeling of the obtained clusters, such as the characterizing topics of each discovered era, thus adding a semantic dimension to our analysis.

1 Introduction

In the last years, much attention has been devoted to topics related to Social Network Analysis. One research direction that has attracted researchers in various fields, including Data Mining, is analyzing networks that evolve over time. Time in networks can play a double role: the entities involved may perform actions, and the connectivity structure may change. In this last setting, several phenomena can be analyzed, and much effort has been devoted in this direction so far [13,12,10,4,3].

In this paper, we focus on detecting clusters of temporal snapshots of an evolving network, to be interpreted as *eras* of evolution of the network. By analyzing the similarity of the structures of consecutive temporal snapshots of the same network, we observe that, despite a global increase of similarity, it is possible to detect periods of sudden change of behavior, where people act in a

M.J. Zaki et al. (Eds.): PAKDD 2010, Part I, LNAI 6118, pp. 81–90, 2010.

counter-trend fashion, making this similarity either decrease, or suddenly start increasing very fast, much more than the average.

In real-life social networks, in fact, a common phenomenon is that people tend to both keep being part of the networks, and keep alive all the connections created in the past. On the other hand, new users join the networks as time goes by, and people set new relationships while keeping the previous ones[13]. However, while the number of newly created relationships tends to be almost constant at every snapshot, the number of previous relationships kept alive grows, thus the global effect of newly added nodes or edges looses importance over time [3]. Because of this, the similarity of the structure of two consecutive temporal snapshots increases almost at each step. The increase, however, is not locally uniform: for example, there can be one snapshot where suddenly people change behavior and start giving more importance to creating new connections, In other words, despite a global moderate *conservative* trend, people can suddenly alternate highly more conservative periods, or a highly more *innovative* behavior.

The aim in this paper is to catch these sudden changes by detecting the snapshots in which they start. Intuitively, these are starting points of new eras. In a globally changing world, we then want to detect eras characterized not by changes in structure (that we not only allow within the same cluster of snapshots, but we also expect), but rather characterized by a change in counter-trend with the previous era: either the previous results more conservative, or it is actually more innovative than the era under investigation.

To this aim, we introduce a novel hierarchical clustering methodology, based on a dissimilarity measure derived from the Jaccard coefficient computed between two temporal snapshots of the network. We devise a framework to discover and browse the era hierarchy either in top-down or a bottom-up fashion, from the lowest level of the single temporal snapshots, to the highest level of the complete period of existence of the network.

In order to do so, we find a measure of the dissimilarity of two temporal snapshots, and we show how to use it as a basis for detecting starting points of new eras. In our experimental section, we show how this measure is not affected by classical phenomena detectable in real-life networks, such the presence of highly connected nodes. We apply this methodology to real data, extracted by the well known bibliographic database DBLP. We build a co-authorship network from it, and analyze two different aspects of the network, namely the nodes (authors), and the edges (collaborations).

Our contribution can be then summarized as follows: we define a dissimilarity measure between two temporal snapshots of an evolving network; we describe the clustering process driven by this measure; we show how to apply labels to the obtained clusters, in order to add a semantic dimension to our analysis; we present the results obtained on the DBLP network.

2 Related Work

Interesting properties have been recently studied and discovered on evolving networks, such as shrinking diameters, and densification power law. Specifically, the

authors in [13] discover that in most of these networks the number of edges grows superlinearly in the number of nodes over time and that the average distance between nodes often shrinks over time. In literature, many models capturing these properties have been proposed; an interesting survey is presented in [7].

Three more recent works are [12,14,16]. In the first, Leskovec et al. present a detailed study of network evolution. They analyze four networks with temporal information about node and edge arrivals and use a methodology based on the maximum-likelihood principle to show that edge locality plays a critical role in evolution of networks. In the second, McGlohon et al. study the evolution of connected components in a network. In [16], the authors propose a novel model which captures the co-evolution of social and affiliation networks.

The notion of *temporal graph* has been studied in [10]. The main aim of this paper is to study how do the basic properties of graphs change over time. A similar setting is used in [11] where Kossinets et al. study the temporal dynamics of communications. They define a temporal notion of "distance" in the underlying social network measuring the minimum time required for information to spread between two nodes. Other works related to the temporal analysis in a network propose the study of aspects of the temporal evolution of the Web [6,8,9,5].

For our temporal analysis we perform hierarchical clustering: an interesting survey on existing clustering approaches can be found in [2].

3 Problem Definition

We are given an evolving network G, whose evolution is described by a temporally ordered sequence of temporal snapshots $T = \{t_1, t_2, \ldots, t_n\}$, where t_i represents the i-th snapshot. T can be either computed on the sets of nodes, i.e. each snapshot t_i is represented by the set of nodes involved, or on the sets of edges, i.e. each snapshot is represented by the set of edges in it.

Based on a dissimilarity measure $d : (t_i, t_{i+1}) \rightarrow]-\infty, +\infty[$, we want to find a hierarchical clustering on T, returning clusters $C_i = \{t_j, \ldots, t_{j+k}\}$, with $j \geq 1$, and $0 \leq k \leq n - j$.

Each cluster represents then an era of evolution. Due to the global evolution of real-life networks, we do allow alterations of the structure of the network among snapshots of the same cluster, as long as they follow a constant trend. As soon as this trend changes, we want to set the corresponding snapshot as the first of a new era. The stronger is the change, the higher should be the dissimilarity of that snapshot with the previous one. The definition of the dissimilarity function should reflect this intuition.

We then want to assign to each cluster C_i a label describing the represented era. This step adds a semantic dimension to our framework.

4 Framework for Temporal Analysis

In this section we describe the details of the framework that we propose.

Dissimilarity. In order to perform clustering, the first step is to define a measure of dissimilarity among elements that we want to cluster. In our setting, a simple way to do this is to use the Jaccard coefficient. In a generic network, we can easily apply this coefficient on either two sets of nodes or two sets of edges, where each set corresponds to a temporal snapshot of the network. As we show later in the paper, clustering temporal snapshots actually corresponds to perform a segmentation of the sequence of the snapshots, thus we are interested only in computing this Jaccard coefficient for every pair of consecutive snapshots.

Real-life networks are well known to follow global evolutionary trends, then if we plot the Jaccard coefficient for each snapshot, we shall see a global increase (or decrease), characterized by an almost constant slope of the Jaccard coefficient plot, alternated by (moderate to high) changes of this slope (we prove this intuition in Section 5). An immediate way, to define starting point of new eras is to detect the snapshots corresponding to these changes. This could be done by computing the second derivative of the Jaccard and finding values different from zero. However, the Jaccard is continuous but not derivable exactly in the points we need. To overcome this problem, we consider an approximation of the second derivative defined as follows. We take triples of consecutive years, and we trace the segment that has as endpoints the Jaccard computed for the first and the third snapshot. If the middle point is distant from the segment, the corresponding snapshot should be considered as the start of a new era. The Euclidean distance between the middle point and the segment also gives as a quantitative analysis of how important is the change: the higher the distance, the higher the change.

Definition 1. *Given a temporal snapshot t_j, we define the following measure:*

$$s_N(t_j) = \frac{|c_N(t_j) - (m \times j) - q|}{\sqrt{(1 + (m^2))}}$$

where $m = \frac{c_N(t_{j-1}) - c_N(t_{j+1})}{t_{j-1} - t_{j+1}}$, $q = (-(j+1) \times m) + c_N(t_{j+1})$, *and* $c_N(t_k) = \frac{|N_{k-1} \cap N_k|}{|N_{k-1} \cup N_k|}$ *is the Jaccard coefficient computed on the node sets.*

Defining s_E, which is the counterpart computed on the set of edges, requires to consider c_E instead of c_N, where c_E is the Jaccard computed on the edges.

However, this measure takes, formally, only one snapshot as input, thus it is not intuitive to use as basis for a clustering methodology. In order to tackle this problem, we define a dissimilarity between any two snapshots as follows.

Definition 2. *Given an ordered sequence $t_1, t_2, \ldots, t_n$ of temporal snapshots of a network G, the dissimilarity between any two snapshots t_i and t_j computed on their node sets is defined as*

$$d_N(t_i, t_j) = \begin{cases} s_N(t_{max(i,j)}) & \text{if } |i - j| = 1 \\ undefined & otherwise \end{cases}$$

Defining the similarity on the edges d_E requires to consider s_E instead of s_N.

Moreover, this dissimilarity measure allows for a straightforward hierarchical clustering: an higher dissimilarity corresponds to a stronger separation between

two consecutive eras. This means that by setting fixed threshold, we can draw a dendrogram of the hierarchical clustering, driven by this dissimilarity as a criterion for merging two consecutive clusters in a bigger one.

Merging Clusters. In hierarchical clustering, when merging clusters, there are various main approaches followed in the literature to define the distance between two clusters: the maximum distance between any two points belonging to the two clusters (complete linkage), the minimum (single linkage), the average (average linkage), the sum of all the intra-cluster variance, and so on.

Given two clusters $C_i = \{t_1, t_2, \ldots, t_k\}$ and $C_j = \{t_{k+1}, t_{k+2}, \ldots, t_{k+p}\}$, in order to define the distance between two clusters, we shall first compute all the distances between every pair (t_i, t_j), with $1 \leq i \leq k$ and $k+1 \leq j \leq k+p$.

However, according to Definition 2, only one pair of snapshots has a dissimilarity defined: (t_k, t_{k+1}). At this point, we use this dissimilarity as inter-cluster distance. As one can immediately see, taking the only available dissimilarity value as distance between clusters actually corresponds not only to both the complete linkage and the single linkage, but also to the average. In our case, thus, the three of them are identical.

Assigning Labels to Clusters. Once we have computed the cluster hierarchy, we want to add a description of every era. In order to do so, we label each cluster with the node (or edge, or a property of it), that maximizes the ratio between its relative frequency in that cluster, and its relative frequency in the entire network. This strategy may produce several values equal to 1 (identical numerators and denominators). In order to discern among these cases, we weight the numerator by multiplying it again for the relative frequency in the cluster under analysis. In this way, we give more importance to 1s deriving from nodes (or edges) with a higher number of occurrences in the cluster.

With this frequency based strategy, we are assigning labels that truly characterize each cluster, as each label is particularly relevant in that cluster, but less relevant for the entire network.

One important caveat in this methodology is what to take as label for the edges. In fact, while for the nodes it is straightforward to consider the identity of the corresponding entity of the network as candidate label, the edge express a relationship with a semantic meaning, thus each network requires some effort in defining exactly which label could be applied to a cluster computed on edges. For example, in a co-authorship network, where two authors are connected by the papers that they have written together, a possible strategy is to take every keyword in the title of the papers as possible label. In the experimental section we show exactly this kind of labeling.

5 Experiments

From the DBLP[1] database, we created a co-authorship graph for the years 1979-2006, where each node represents an author and each edge a paper written

[1] http://dblp.uni-trier.de

together by the two connected authors. We then considered each year as temporal snapshot of DBLP, generating then 28 snapshots. In each snapshot we put only the nodes or the edges appearing in the corresponding year, thus not following a cumulative approach.

Jaccard Coefficient. Figure 1(a) shows the Jaccard on both the nodes and the edges. These plots confirm the general increasing behavior of the Jaccard during time, both on nodes and on edges, broken by short series of years in which people acted in counter-trend. Two questions might be raised on the effectiveness of following a Jaccard-based approach for clustering eras: what would the Jaccard computed on non consecutive snapshot tell us? Is the Jaccard noise free?

We start answering the first question by plotting the coefficient computed for every pair of snapshots: figures 2(a) and 2(b) show that the Jaccard decreases when computed between snapshots more distant in time. This observation justifies a dissimilarity measure that takes into account only consecutive snapshots, as two distant snapshots are not likely to be similar, thus they will belong to different clusters. Temporal segmentation is then a good model for clustering real-life evolving networks. Please note that while in this paper we only show the results obtained on one dataset, these considerations are well accepted in the literature regarding evolving networks [3,13].

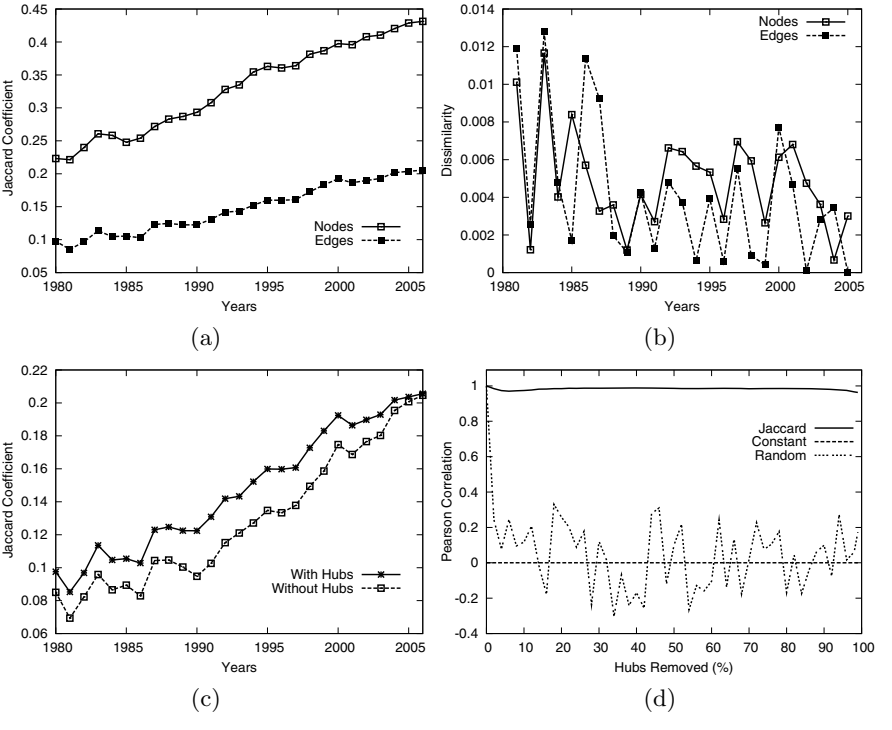
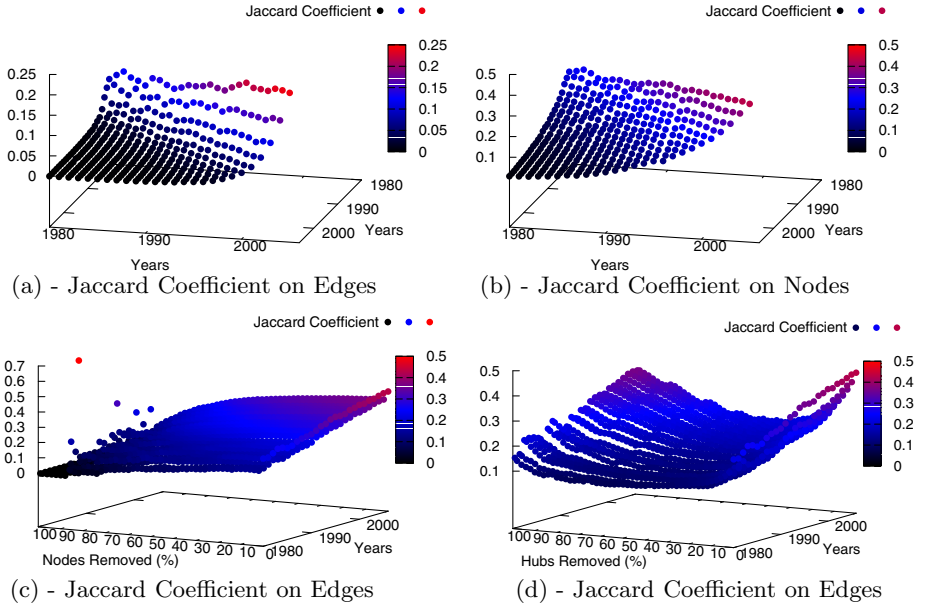


Fig. 1. (a) Evolution of the Jaccard Coefficient in DBLP; (b) Dissimilarity in DBLP; (c) Jaccard Coefficient with and without hubs; (d) Pearson Correlation



(a) - Jaccard Coefficient on Edges
(b) - Jaccard Coefficient on Nodes
(c) - Jaccard Coefficient on Edges
(d) - Jaccard Coefficient on Edges

Fig. 2. Jaccard Coefficient in DBLP

Note that by answering the second question we also say something about possible noise on the dissimilarity measure. As the two are closely related, proving that the Jaccard is noise free also proves the same for the dissimilarity. However, using the Jaccard to this aim results much more intuitive. To answer the second question, we analyzed what happens to the Jaccard coefficient when removing a possible cause of noise in the structure: from the entire network, we removed the 1% of top connected hubs, i.e. highly connected nodes, and recomputed the Jaccard. Figure 1(c) shows that, despite a general decrease of the Jaccard values, the global increasing trend, as well as the local sudden changes, are almost unchanged. In order to further support this observation, we plotted the Jaccard coefficient calculated on different versions of the network snapshots after an increasing percentage of hubs removed. Figure 2(d) shows an interesting result: while the global Jaccard dramatically decreases after removing about 10% of the top hubs, always keeping the global evolutionary behavior, it increases again after removing 70-80% of the hubs. This behavior can be explained by considering the intrinsic inter-components function of hub nodes: after removing the majority of the hubs, we have only the small connected components left in the network, and each of them keeps a high Jaccard during its evolution, acting as a separated network. This does not happen when removing an increasing percentage of random nodes (Figure 2(c)), which makes the Jaccard index globally decrease. As last proof of the strength of the Jaccard as good similarity measure, we report in Figure 1(d) the values of the Pearson correlation [1] between the series of Jaccard coefficient computed on the original dataset, and the ones

computed after removing the hubs. The figure shows that removing an arbitrary percentage of hubs does not affect the correlation with the Jaccard computed on the original network. In the figure we have also reported the correlation with a constant series and a random one.

Dissimilarity. The second step of the framework required to compute our dissimilarity on the basis of the Jaccard coefficient computed on the network. Figure 1(b) reports the values of the dissimilarity for both the edge and the node cases. As one can see, the quantitative analysis of our dissimilarity measure is effective: its values have a considerable standard deviation. That is, we can effectively perform hierarchical clustering finding a well distributed strength of starting snapshots for new eras of evolution.

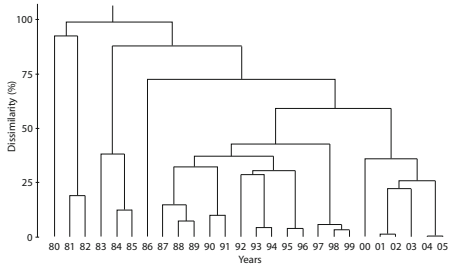
Another observation that can be done is that while the Jaccard values computed on nodes or edges look similar, stronger differences can be found in the dissimilarity plots. That is, we expect the eras computed on nodes to slightly differ from the ones computed on the edges.

As last note, we see that in the first years under investigation there are a few very high peaks of dissimilarity. This is mainly due to the data acquired by DBLP before year 1990. In the first decades, in fact, the set of publications recorded in the database was more restrictive, and sometimes limited only to publications in German. This created a kind of noise in the cluster, and it is the reason why in the final dendrograms we see the years up to 1985 to be among the last ones to be added to the global cluster (Figure 3(a)-(b)), and why also we see labels in German in the final labeling (Figure 3(c)).

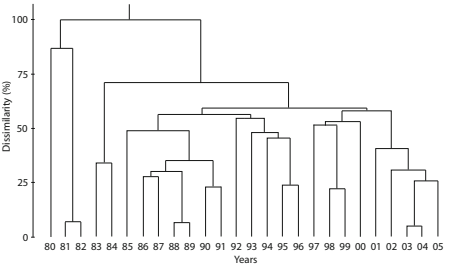
Merging Clusters. We then started to compute the clusters on the sequences of temporal snapshots. We started from clusters containing only one year and then, driven by the dissimilarity values computed in the previous step, we merged similar consecutive clusters, with increasing values of dissimilarity. Figure 3(a)-(b) show the dendrograms of clusters obtained using the sets of edges (on the left) or the sets of nodes (on the right). Please note that the dissimilarity is reported in percentage of the highest value found.

As one can see, there are actually a few differences between the two dendrograms, even if at a higher level the two look similar. A deeper view would show a more uniform distribution of the joins between clusters. We recall that in the dendrogram on the left we report the analysis performed on the edges, i.e. the collaborations, while on the right we show the clusters obtained on the nodes, i.e. the authors. This can be read saying that, while there are different, uniformly distributed, ways of changing evolutionary behavior for the collaboration, there are less thresholds for changing eras while looking at the nodes: i.e., the strength of changes of eras can be further clustered in a few similar thresholds.

Assigning Labels. As last step in our framework, we computed the labels for each cluster obtained. Figure 3 reports, in (c) and (d), 5 labels for each cluster with at least two years (due to space constraints, we do not report labels for single years). Please note that the rows in these tables are sorted by order of



(a) - Edge Eras



(b) - Node Eras

ERA	TOPICS
04-05	servic, web, mobil, detect, wireless
01-02	web, mobil, softwar, adapt, dynam
98-99	object, parallel, architectur, simul, softwar
95-96	parallel, databas, simul, abstract, logic
93-94	parallel, logic, objectori, databas, abstract
97-99	parallel, object, databas, sotwar, environ
88-89	logic, parallel, expert, databas, languag
90-91	parallel, logic, abstract, languag, databas
84-85	prolog, expert, databas, abstract, languag
87-89	parallel, logic, expert, prolog, databas
81-82	comment, pascal, languag, microcomput, databas
01-03	web, mobil, servic, softwar, architectur
01-05	web, servic, mobil, dynam, detect
92-94	parallel, logic, objectori, databas, languag
92-96	parallel, logic, databas, abstract, objectori
87-91	parallel, logic, abstract, databas, languag
00-05	web, mobil, servic, dynam, adapt
87-96	parallel, logic, databas, abstract, languag
83-85	databas, prolog, abstract, expert, languag
87-99	parallel, logic, databas, abstract, languag
87-05	parallel, architectur, softwar, perform, environ
86-05	parallel, architectur, softwar, perform, environ
83-05	parallel, architectur, softwar, perform, gener
80-82	pascal, languag, rechnungswesen, comment, databas
80-05	parallel, architectur, softwar, perform, gener

(c) - Edge Era Labels

ERA	AUTHORS
03-04	Z. Wu, W.Y. Ma, H. Zhang, M. Li, W. Gao
88-89	M. Ali, C. Lécluse, C. Tong, A.J. Brodersen, P. Richard
81- 82	W. Lipski Jr., M. Courvoisier, A.C. Klug, N. Goodman, R. Keil
98-99	M. Potkonjak, G. J. Edwards, R. Vemuri, P. Kuosmanen, R. Radhakrishnan
90-91	V. Zue, R. Potasman, A. R. Newton, M.S. Phillips, A. Srinivasan
95-96	H.J. Nussbaumer, T.C. Fogarty, R.M. Owens, R. Yagel, M. Tien-Chien Lee
03-05	W.Y. Ma, Zhaohui Wu, Minglu Li, Licheng Jiao, Wen Gao
86-87	S.Khoshafian, R.F. Rashid, J. van de Graaf, D.S. Lindsay, K. Doshi
86-89	S. Khoshafian, C. Lécluse, M. Ali, J.M. Bower, G.P. Copeland
02-05	W.Y. Ma, M.T. Kandemir, Z. Wu, W. Gao, M. Li
83-84	H. Bekic, G.Spur, W. Frey, F.L. Krause, W. Remmele
86-91	W. Ameling, D. Chaum, G.E. Kaiser, E.Y. Shapiro, C. Lcluse
01-05	M.T. Kandemir, W. Gao, H. Zhang, W.Y. Ma, Z. Wu
94-96	T.C. Fogarty, B. Kaminska, B. Lin, R.M. Owens, R.K. Brayton
93-96	R.K. Brayton, B. Kaminska, T.C. Fogarty, A.L. Sangiovanni-Vincentelli, R.L. Wainwright
85-91	W. Ameling, D. Chaum, E.Y. Shapiro, D.W. Stemple, S. Khoshafian
97-99	M. Potkonjak, A.N. Choudhary, B. Schneier, C.J. Taylor, A.Y. Levy
97-00	M. Potkonjak, T.S. Huang, A.N. Choudhary, C.J. Taylor, M.M. Veloso
97-05	M.T. Kandemir, E.R. Hancock, W. Gao, H. Zhang, T.S. Huang
92-96	R.K. Brayton, A.L. Sangiovanni-Vincentelli, J.D. Foley, B. Kaminska, D.K. Pradhan
85-96	A.L. Sangiovanni-Vincentelli, R.K. Brayton, M. Sharir, K. Kennedy, M.J. Carey
85-05	E.R. Hancock, A.L. Sangiovanni-Vincentelli, P.S. Yu, S.M. Reddy, M.T. Kandemir
83-05	E.R. Hancock, A.L. Sangiovanni-Vincentelli, S.M. Reddy, P.S. Yu, M.T. Kandemir
80-82	P. Raulefs, N. Goodman, S. Kartashev, S. Kartashev, G. Winterstein
80-05	E.R. Hancock, A.L. Sangiovanni-Vincentelli, S.M. Reddy, P.S. Yu, M.T. Kandemir

(d) - Node Era Labels

Fig. 3. DBLP Eras discovered on Edges or Nodes

cluster formation. Please also note that the keywords of the publications were pre-processed using the Porter's stemming algorithm [15].

We recall that for each cluster C_i we assign the set of the k labels maximizing the ratio between their frequency in C_i and their frequency in the entire network. Due to our strategy, it is then not surprising that, if we look at the node cluster labels, in all the clusters except the complete network we do not find the most active authors, but the ones that mostly published only on each specific cluster. That is, we can find as labels authors with a not so strong publication record, but whose publication record was extremely stronger in a specific cluster w.r.t the entire network. This behavior is less evident in the edge era labels, where topics such as "parallel" can be found in different clusters.

In this table, however, another consideration can be done. If we compare the era labels with the dissimilarity plot, we can see which are the labels that correspond to more conservative or more dynamic eras. If we exclude the first noisy years, the highest peak in the dissimilarity plot is around year 2000. This, in fact, corresponds to the creation of the cluster starting at year 2001. We can say that from year 2000, a short era of very conservative collaborations started. One of the most representing label for the collaborations in these cluster is "web". One can say that this topic is highly representative for highly conservative collaborations, i.e., collaborations that take place among the same (large) group of people.

6 Conclusions and Future Work

We have proposed a framework for the discovery of eras in an evolving social network. Based on a dissimilarity measure derived from the Jaccard coefficient, we have presented a methodology to perform hierarchical clustering of the temporal snapshots of a network. We have applied our methodology to real-life data, showing the effectiveness of our approach.

Future research directions include the application to several different evolving networks, possibly showing different temporal behaviors, different definitions of dissimilarity, and the introduction of a label-driven temporal clustering strategy.

References

1. Thirteen ways to look at the correlation coefficient. The American Statistician 42(1), 59–66 (1988)
2. Berkhin, P.: Survey of clustering data mining techniques. Technical report, Accrue Software, San Jose, CA (2002)
3. Berlingerio, M., Bonchi, F., Bringmann, B., Gionis, A.: Mining graph evolution rules. In: ECML/PKDD, vol. (1), pp. 115–130 (2009)
4. Berlingerio, M., Coscia, M., Giannotti, F.: Mining the temporal dimension of the information propagation. In: Adams, N.M., Robardet, C., Siebes, A., Boulicaut, J.-F. (eds.) IDA 2009. LNCS, vol. 5772, pp. 237–248. Springer, Heidelberg (2009)
5. Bordino, I., Boldi, P., Donato, D., Santini, M., Vigna, S.: Temporal evolution of the uk web. In: ICDM Workshops, pp. 909–918 (2008)
6. Brewington, B.E., Cybenko, G.: Keeping up with the changing web. IEEE Computer 33(5) (2000)
7. Chakrabarti, D., Faloutsos, C.: Graph mining: Laws, generators, and algorithms. ACM Comput. Surv. 38(1) (2006)
8. Cho, J., Garcia-Molina, H.: Estimating frequency of change. ACM Trans. Internet Techn. 3(3), 256–290 (2003)
9. Gomes, D., Silva, M.J.: Modelling information persistence on the web. In: ICWE 2006: Proceedings of the 6th international conference on Web engineering, pp. 193–200 (2006)
10. Kempe, D., Kleinberg, J.M., Kumar, A.: Connectivity and inference problems for temporal networks. In: STOC, pp. 504–513 (2000)
11. Kossinets, G., Kleinberg, J.M., Watts, D.J.: The structure of information pathways in a social communication network. In: KDD, pp. 435–443 (2008)
12. Leskovec, J., Backstrom, L., Kumar, R., Tomkins, A.: Microscopic evolution of social networks. In: KDD, pp. 462–470 (2008)
13. Leskovec, J., Kleinberg, J.M., Faloutsos, C.: Graphs over time: densification laws, shrinking diameters and possible explanations. In: KDD, pp. 177–187 (2005)
14. McGlohon, M., Akoglu, L., Faloutsos, C.: Weighted graphs and disconnected components: patterns and a generator. In: KDD, pp. 524–532 (2008)
15. Robertson, S.E., van Rijsbergen, C.J., Porter, M.F.: Probabilistic models of indexing and searching. In: SIGIR, pp. 35–56 (1980)
16. Zheleva, E., Sharara, H., Getoor, L.: Co-evolution of social and affiliation networks. In: KDD, pp. 1007–1016 (2009)