Discovering Eras in Evolving Social Networks

Michele Berlingerio, Michele Coscia, Fosca Giannotti,
Anna Monreale, Dino Pedreschi

ISTI - CNR, Pisa, Italy
Computer Science Dep., University of Pisa, Italy

{michele.berlingerio,fosca.giannotti}@isti.cnr.it,
{coscia,annam,pedre}@di.unipi.it

PaKDD 2010, Hyderabad

Background

- Increasing interest in analyzing evolving networks
- Time in networks: two possible scenarios
  - Action: entities do something
  - Evolution: change in the structure
- The two may coexist!
- Example: online social networks

Motivation

- How does a network evolve over time?
- Is the evolution somehow regular?
- Is the evolution characterized by important *eras*?
- How do we find such eras and how can we discern among them?

Previous Works

Analysis and Mining both at global and local level

- Leskovec et al [KDD05], evolution of global properties
- McGlohon et al [KDD08], evolution of connected components
- Berlingerio et al [PKDD09], mining graph evolution rules

Framework for Discovery of Eras

- Extraction of a time evolving network from real data
- Definition of a measure of dissimilarity among temporal snapshots of the same data
- Definition of clusters giving thresholds of such dissimilarity
- Merge of two (consecutive) clusters
- Assigning labels to clusters
- Realization of a dendrogram summarizing the clusters

Framework for Discovery of Eras

- Extraction of a time evolving network from real data
- Definition of a measure of dissimilarity among temporal snapshots of the same data
- Definition of clusters giving thresholds of such dissimilarity
- Merge of two (consecutive) clusters
- Assigning labels to clusters
- Realization of a dendrogram summarizing the clusters

Network Extraction

From the DBLP bibliographic database

- Extracted a co-authoship graph from the years 1979-2006
- Each year as different temporal snapshot (not cumulative)
- Edges not weighted

Framework for Discovery of Eras

- Extraction of a time evolving network from real data
- Definition of a measure of dissimilarity among temporal snapshots of the same data
- Definition of clusters giving thresholds of such dissimilarity
- Merge of two (consecutive) clusters
- Assigning labels to clusters
- Realization of a dendrogram summarizing the clusters

Dissimilarity measure

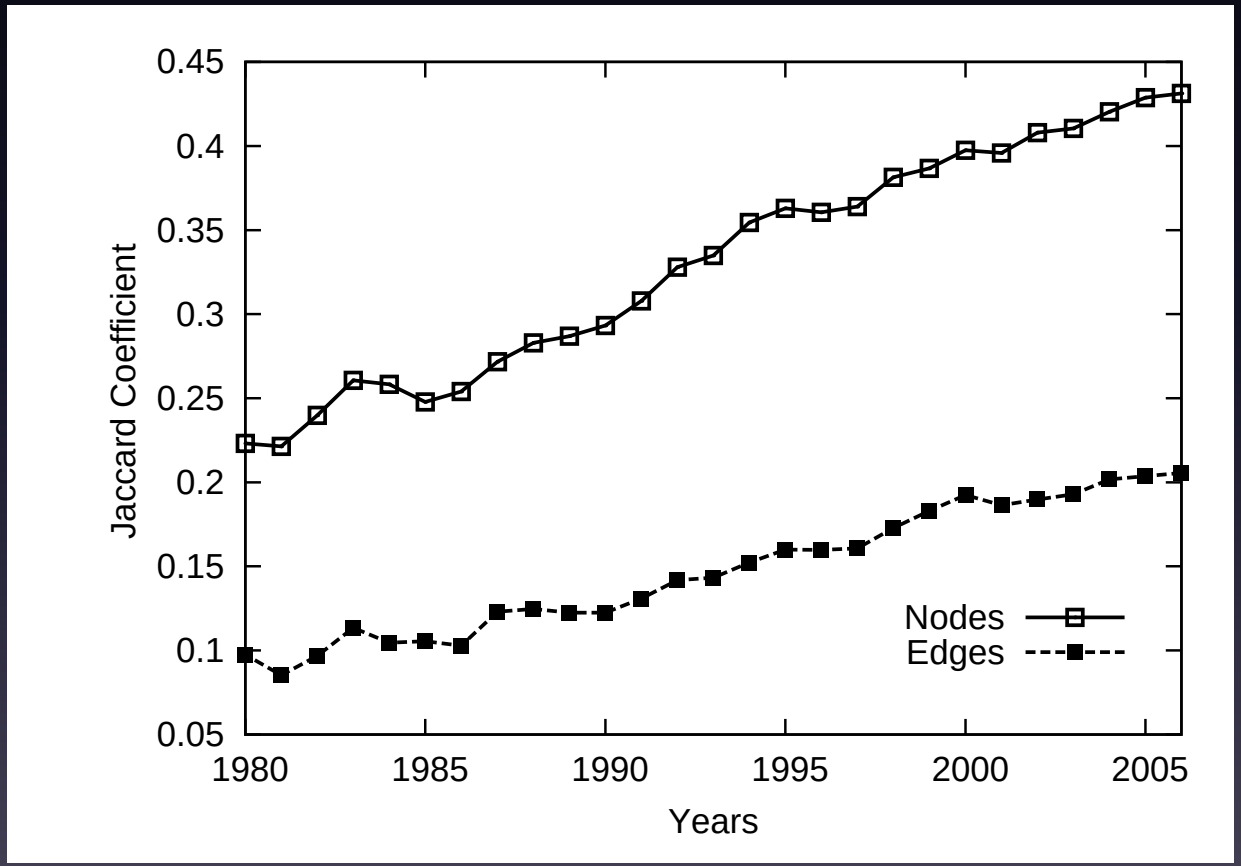
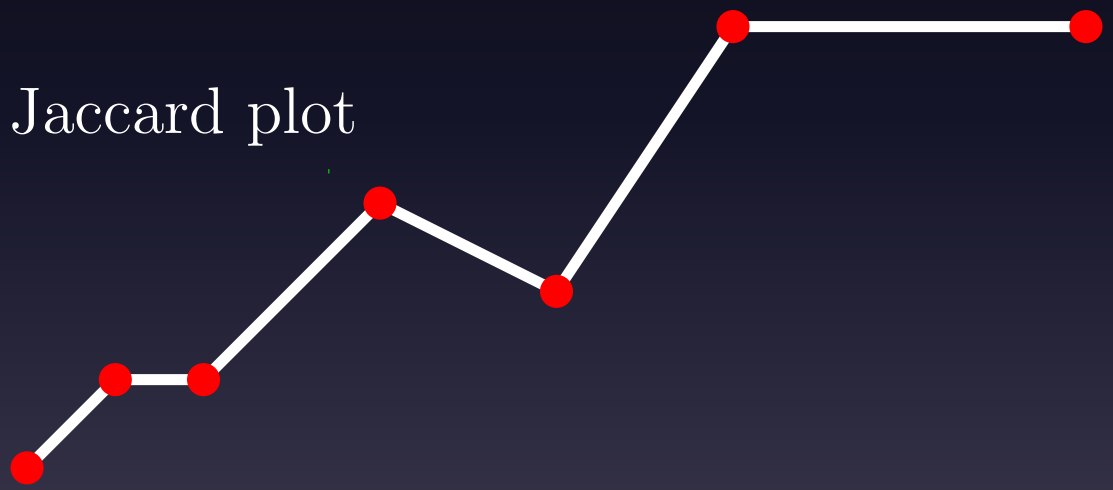
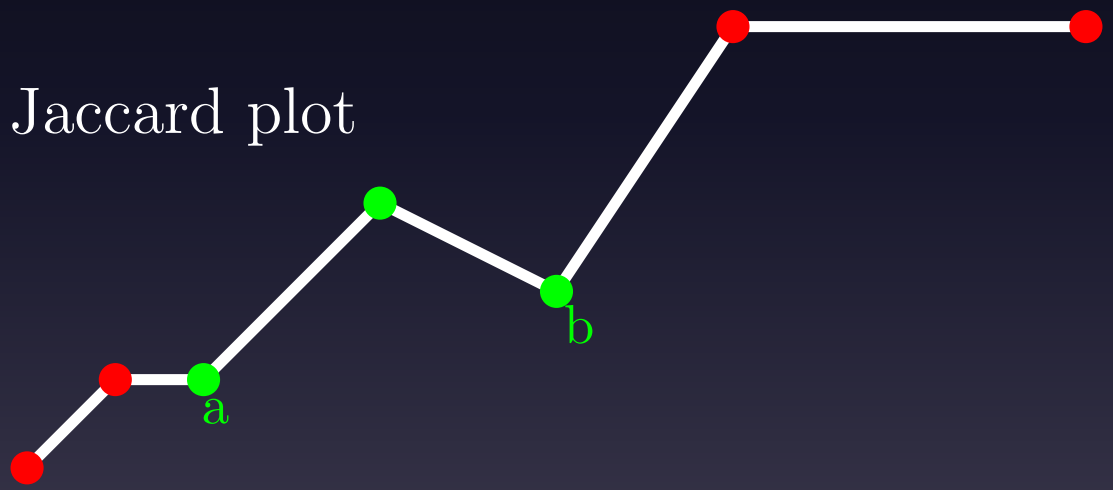


Figure: Evolution of the Jaccard Coefficient in DBLP

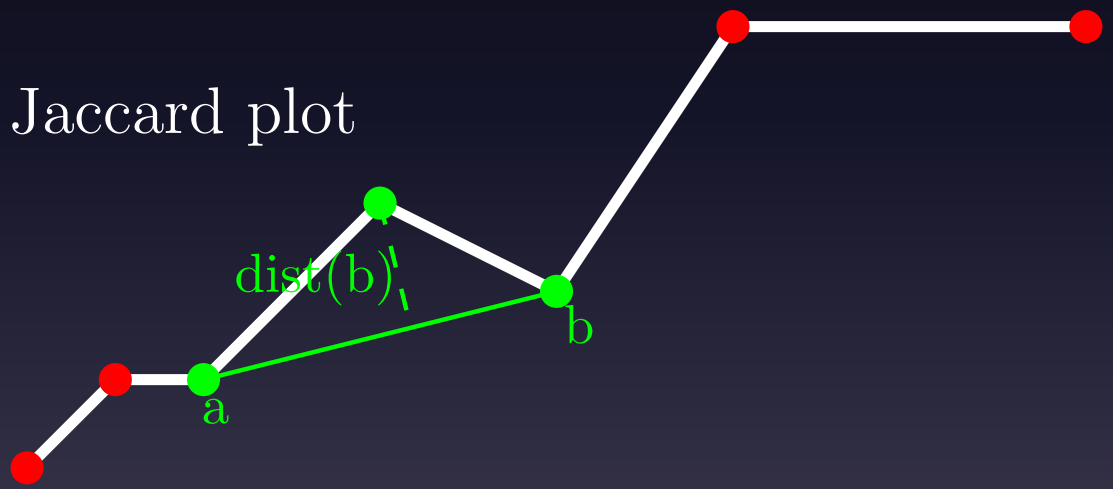
Dissimilarity measure



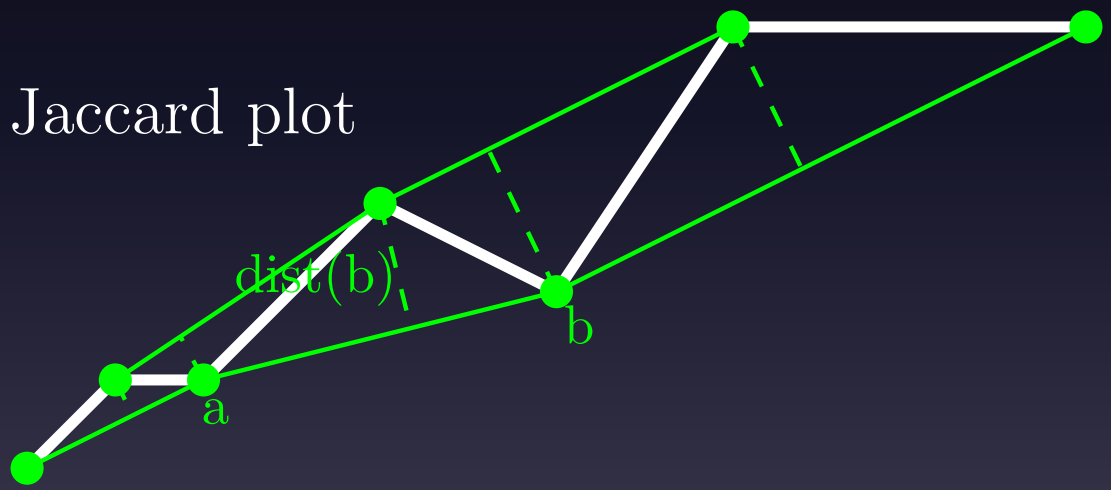
Dissimilarity measure



Dissimilarity measure



Dissimilarity measure



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

Dissimilarity measure

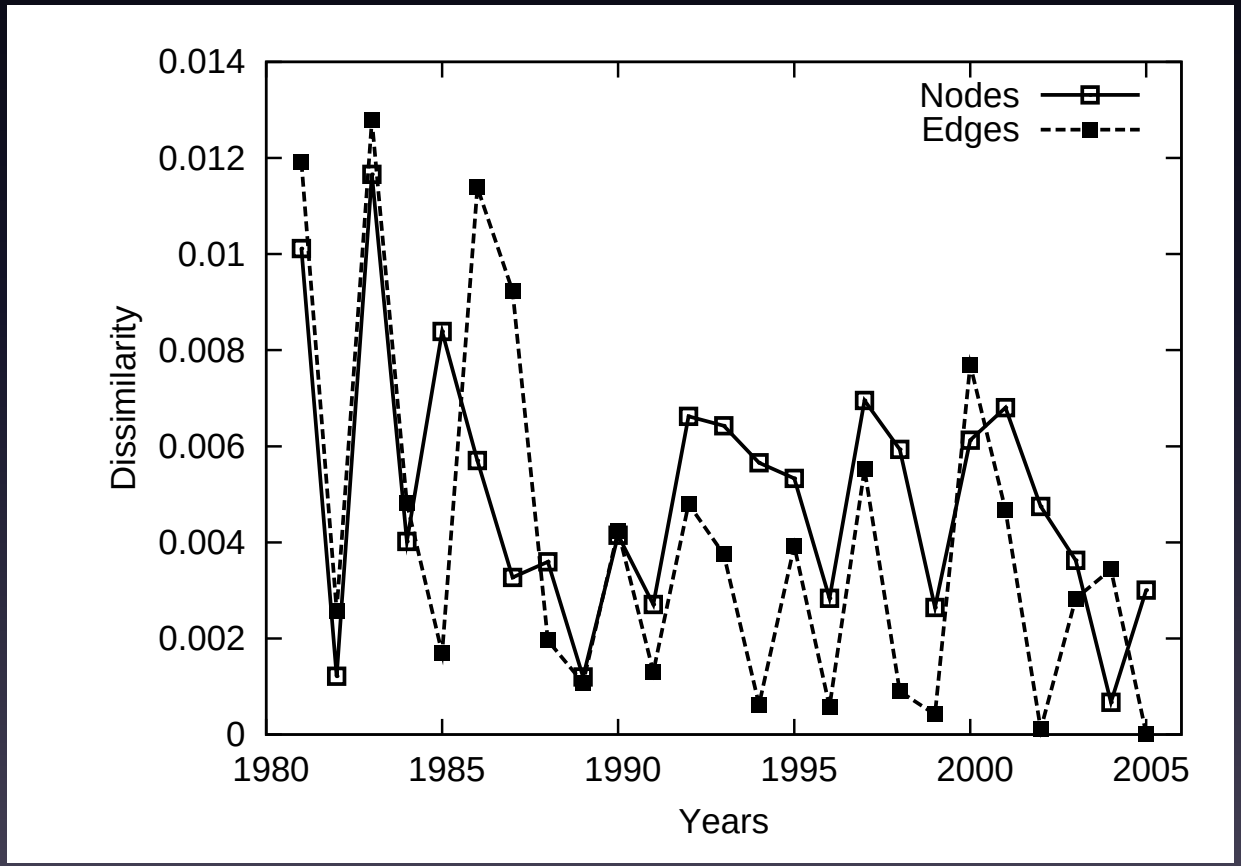


Figure: Dissimilarity Measure in DBLP

Framework for Discovery of Eras

- Extraction of a time evolving network from real data
- Definition of a measure of dissimilarity among temporal snapshots of the same data
- Definition of clusters giving thresholds of such dissimilarity
- Merge of two (consecutive) clusters
- Assigning labels to clusters
- Realization of a dendrogram summarizing the clusters

Defining and Merging Clusters

- Based on our Dissimilarity
- Higher Dissimilarity, stronger separation of cluster
- When merging, we only look at the distance between the endpoints
- Complete, single and average linking coincide

Framework for Discovery of Eras

- Extraction of a time evolving network from real data
- Definition of a measure of dissimilarity among temporal snapshots of the same data
- Definition of clusters giving thresholds of such dissimilarity
- Merge of two (consecutive) clusters
- Assigning labels to clusters
- Realization of a dendrogram summarizing the clusters

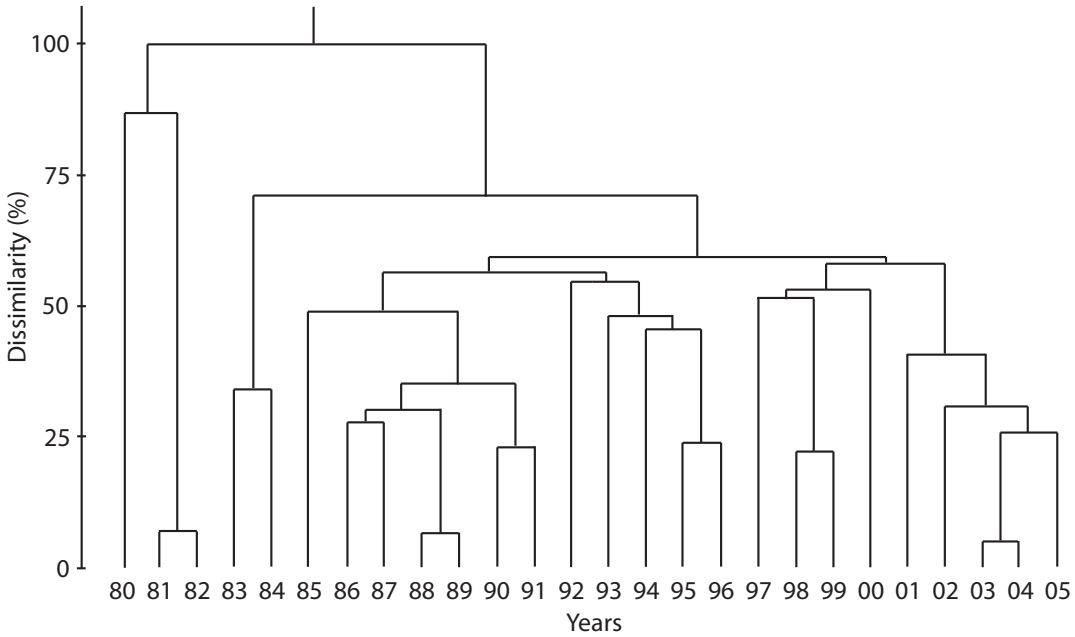
Assigning Labels to Clusters

- Goal: adding a description to every era
- For every era, take the node (or edge) that is frequent in that cluster, but not frequent in all the others
- Trivial for nodes: identities of the nodes as labels
- Edges express a semantic relationship, label depends on the semantic
  - Co-authorship network: title keywords as edge labels

Framework for Discovery of Eras

- Extraction of a time evolving network from real data
- Definition of a measure of dissimilarity among temporal snapshots of the same data
- Definition of clusters giving thresholds of such dissimilarity
- Merge of two (consecutive) clusters
- Assigning labels to clusters
- Realization of a dendrogram summarizing the clusters

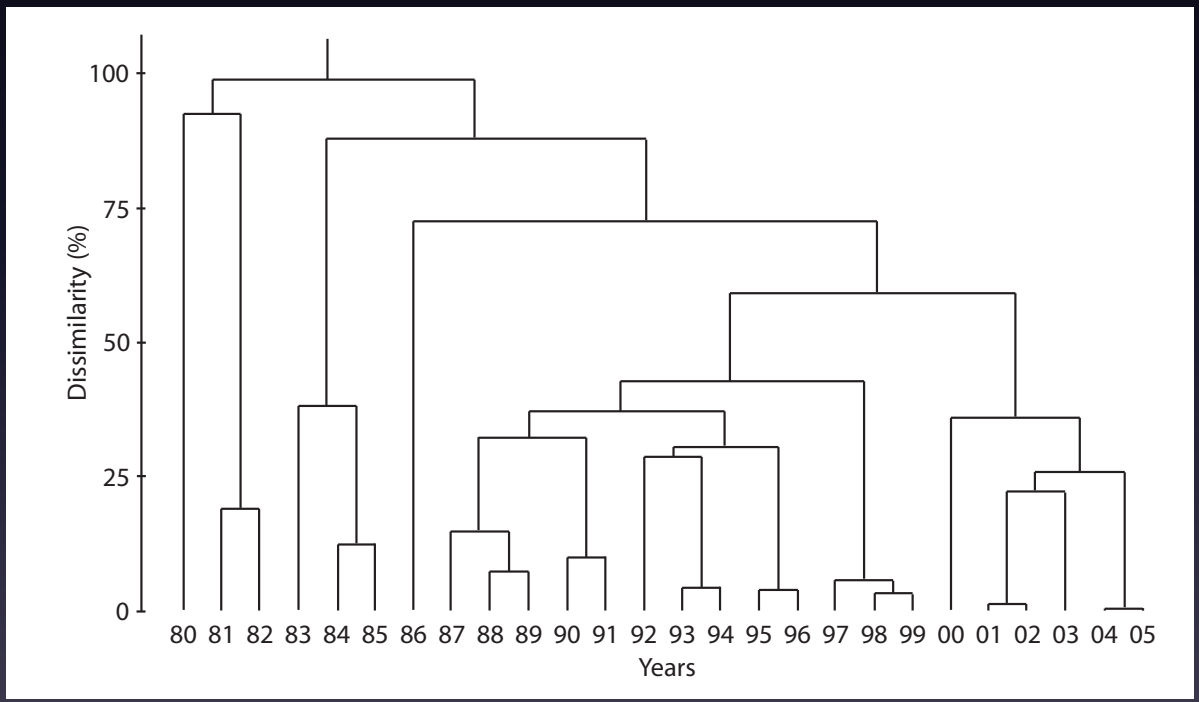
Eras on nodes



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

Figure: Eras on nodes in DBLP

Eras on edges



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

Figure: Eras on edges in DBLP

Conclusions

- Framework for discovering eras in evolving social networks
- Jaccard based dissimilarity measure
- Labels to characterize eras

Future directions:

- Combined node-edge analysis
- Analysis on other datasets
- Definition of a quality measure

Thank you!

Questions?

michele.berlingerio@isti.cnr.it