Prominence in Networks: A Co-evolving Process

Yang Yang & Nitesh V. Chawla
Department of Computer Science and iCeNSA
The University of Notre-Dame
Notre Dame, Indiana
Email: {yyang1,nchawla}@nd.edu

Xiaohui Lu & Sibel Adalı
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, New York
Email: {lux, sibel}@cs.rpi.edu

***Abstract*—We investigate how people and objects that they create (artifacts) gain prominence in collaborative networks. As an example, consider academic research communities where people and their artifacts (research papers) both have prominence. But, these prominence values are linked to each other and evolve together. In particular, for an author to make an impact on the scientific community, she has to interact with diverse sets of individuals. The results of this is that her research will be known by a large number of communities. However, the research communities and topics have to be aligned along her research interests as well. Peer reviewers have to know and understand the research field and the results must be disseminated to the larger community through sustained interest in the given research field. Hence, the prominence of individuals and their artifacts evolve simultaneously. In this paper, we develop novel methods to study both types of evolution processes and show their effectiveness using the DBLP dataset.**

I. Introduction

Studying prominence of individuals or objects created by them (artifacts) in networks is of crucial importance in many networks that rely on efficient methods to locate important information, and filter out less important and noisy values. In particular, ranking information is a crucial function of many information systems. These ranking algorithms must take social relations into account to improve their performance. In social networks, finding the correct individuals in a specific field is crucial.

We consider the author and papers network, where the authors create a social network via collaborations or citations, and the papers are the created objects, which are the artifacts. We posit that in social collaborative networks, prominence of individuals and their artifacts are intricately linked through a co-evolving process. Individuals' prominence is linked to the impact they make by their artifacts. These artifacts need to be acknowledged by their peers. The larger a community they reach, the bigger their impact will be. However, the impact of this will not be high in any single community if that community is not aligned with the specific research problem. For example, a high impact paper in a very new research area may not gain much traction unless others continuously develop this research area and publish in it. The sustained interest in a research area makes it possible for the research area to have an impact, which impacts both the papers and the authors in that area.

A. Contribution

We consider the co-evolving process for prominence as follows. Co-evolution is defined as a process where the nodes or the individuals might affect the network, and the network might affect the nodes or the individuals, as well. To that end, first, we study how the prominence of individuals is predicted by the research groups they belong to and the papers they publish. We consider communities of individuals based on their collaborations, and communities of artifacts, i.e. research areas. We show that individuals who belong to communities that publish in diverse research areas tend to have high prominence.

Next, we study how the research areas evolve over time. Papers that belong to research areas that have attracted sustained interest over the years tend to have higher impact in the research field in general. These research areas show that researchers have come together over time to promote a specific approach to solving problems, and papers that have started this type of movements tend to have high impact.

We show examples from DBLP to illustrate the utility of our approach and discuss how both methods can be combined to provide a better understanding of evolution in social collaborative networks.

II. Basics

We assume a network is composed of actors (people) and artifacts that they collaborate on. In the case of DBLP, actors correspond to researchers or authors, and artifacts are the papers they write. Using this network, we construct an author to author network that has authors as nodes. Two authors are connected to each other if they have a common paper. The weight of each edge is given by the number of common authors. We cluster this graph using a community discovery algorithm, Walktrap [1], to find authors who work together, research collaborators.

Most work on analyzing collaboration networks concentrate on understanding which individuals are linked to each other through social relationships and the specific advantages provided by the network position of the individuals. However in this type of analysis, the rich network of collaborations is reduced to a weight score between individuals. In our work, we also study a parallel artifact to artifact network. This network has artifacts, i.e. papers as nodes. Two artifacts are linked if they share a common author. We compute the weight

978-1-4799-0203-3/13/$31.00 ©2013 IEEE

between two artifacts using the Jaccard measure [2], number of common authors between the two papers divided by the total number of authors in the two papers.

We cluster the paper-to-paper network using a community discovery algorithm such as Walktrap [1] similar to the author to author graph. Each community in this case corresponds to a research area. However, this definition of research area is more fine tuned than publication venues for the following reasons: a venue may have multiple research topics or it may be a very narrow research area. Furthermore, venues may have a range of publications, from very high quality to ones with lower quality. However, a cluster is likely to have objects of similar value as we have shown in our previous work. In fact we show [3] that these clusters are especially useful in finding prominent people, more so than actual venues. In other words, the clusters capture a specific group of collaborators working on a specific research area. These areas may be small and tight as in the case of neural networks for example, or large and more loosely connected as in systems areas. As a result, we will refer to object clusters in this paper as research collaborations in a specific area, or a research area for short.

III. Related Work

It is widely believed that real world networks evolve over time [4]. In the process of network evolution, actors may leave the network, re-construct connections with others, and new actors may join in the network. Hence, the status of actors evolve gradually in these processes, so as the communities in the networks. In [5], Leskovec et al. show that networks tend to become denser over time and the distances tend to get smaller. To explain this behavior, the authors introduce two local processes called community guided attachment and forest fire models that explain how local behavior based on copying links leads to the network level evolution. Backstrom et al. [6] examine the structural features that influence the individuals to choose groups to join in a series of network snapshots. In particular, using DBLP as a testbed, authors use conferences as a stand in for communities. By using decision trees and features based on the individual's the number of friends and the features of the community, they predict whether an individual will join a community in the future and whether the community will grow in the future. In this paper, we study a different problem: how is the prominence of individuals predicted by the network evolution.

Generally, communities in consecutive snapshots are treated as standalone groups. Chakrabarti et al. [7] argue that communities should not change dramatically between consecutive timesteps. They define evolutionary clustering as a task that is faithful to the current network, and consistent with earlier snapshots. Based on this definition, they propose evolutionary settings for k-means and agglomerative hierarchical clustering algorithms.

Most of the existing work focuses on one-mode networks which exclusively consider one specific type of objects and ignore the relationships to other types of objects. In [8], Tang et al. investigate community evolution in dynamic multi-mode networks which are constituted by multiple social actors, with the relationship evolve step by step. However, we consider co-evolution of people communities and objects (created by people) communities, but also take the relationships between people and artifacts into consideration. From these two points, we are trying to show how people become prominent, and this depends on the other people they collaborate with, the communities that they belong to and the evolution of these communities.

A number of papers have analyzed author prominence in DBLP. Co-authorship relationships have been used to infer prominence using centrality type measures [9]. However, this type of analysis throws away a lot of useful data. In [10], Sun et al. argues that ranking actors globally in large networks is not meaningful. Actors have prominence in specific communities, which can be found using a clustering algorithm. NetClus [11] extends this idea to three-types networks with star schema. These algorithms both find clusters based on existing venues and rank each author within a specific research area. In contrast, Adalı et. al. [3] find global prominence of authors based on all their participation in research areas. These areas are found without relying on the existing venues. The object to object communities we study in this paper are analogous to this approach.

IV. Data

The DBLP dataset we used in this paper ranges from 1974 to 2006, thus we split the whole dataset into snapshots and each of them has 5 years span. Two types of networks can be constructed from these snapshots, one is authors collaboration network, and another is papers co-author network. In the authors collaboration networks, the nodes are authors, an edge exists between two authors if they collaborate on common papers, the edge weight is the number of common papers. While for the paper co-author networks, nodes are papers, an edge exists if two papers have common authors, the edge weight is the *Jaccard Coefficient* of common authors. From the network structure properties, such as cluster coefficient and centrality, we can capture the underlying model for networks evolution. For each snapshot, there is an author collaboration network and a paper co-author network, thus there are clusters for each type of network. By the paper-author relationship, we can construct the mapping between these two networks' clusters (Figure 1). For example, an authors cluster in author collaboration network can be mapped several papers clusters in paper co-author network, if papers are written these authors. We use the Walktrap algorithm [1] for cluster analysis, as it is non-parametric, effective for large networks, and provides non-overlapping clusters. Note that in the construction of authors' collaboration network, we remove authors who have less than 5 papers, and for papers co-author network, papers written by more than 10 authors or by only one author are also removed.

V. Network Evolution Analysis: Individual-based

The rise of an individual in social collaborative networks is woven with co-evolution processes — the individuals' impact

Fig. 1. Mappings Between Author and Paper

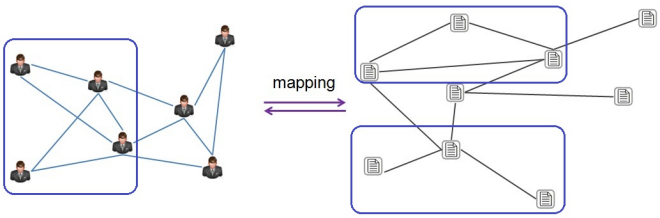
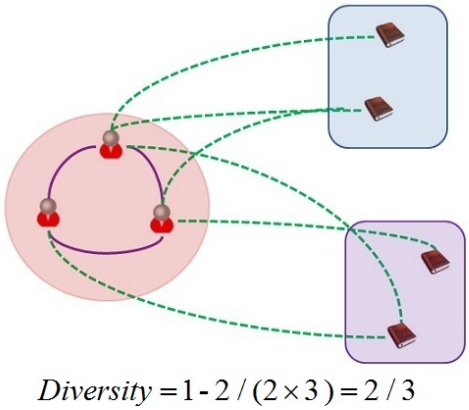


Fig. 2. Diversity Toy Example



is driven by the objects or artifacts (papers in DBLP) that they create as well as from the acknowledgement of the paper from the peers. The community affects the impact of the paper, implying the prominence of the individual.

The questions of interest to our work are: what is the influence of authors and their research artifacts (papers) vis-a-vis the community or group that the individual belongs to? The community of an individual is based on collaborations. In this section, we consider the evolution from the individual (author) perspective and the paper perspective. We first define the notion of *diversity* that captures the notion of popularity of a research topic, and then present analysis on the impact of an author or a paper on the community evolution.

A. Diversity

As scholarship in an academic community evolves, it leads to creation of multiple and diverse research topics and communities. Moreover, an active community can also lead to a number of sub-topics; for instance, within the topic of data mining there might be a sub-topic of classification or social networks or unsupervised learning. Of course, there will also be a potential overlap among the communities, but one can conjure the notion of diversity of not only sub-groups, but also authors' memberships. To some extent the diversity of research topics within a large academic community can represent the popularity of this community. We introduce a metric called *diversity* to measure this.

Authors' cluster in the authors' collaboration network may be related to several different clusters in the papers network (based on the paper-author relationship). This mapping or relationship can be defined by diversity: how many of the authors in the authors network map correctly to clusters in the paper network?

$$diversity(C) = 1 - \frac{a}{a+b} \tag{1}$$

where a is the number of pairs that are correctly clustered in the mapped clusters (C), b is the number of pairs that are mis-clustered in the mapped clusters.

For example for a cluster of authors in authors collaboration network, there will be a set of their publications $\mathbf{P}$, based on the paper-author relationship. This set of papers might be clustered in to several different subsets in the paper network. (see Figure 2). If all of publications in $\mathbf{P}$ are clustered together, then $diversity = 0$, if they are partitioned into large number of subsets, the $diversity$ will be very large. Figure 2 gives an toy example for the calculation of $diversity$. Thus, diversity allows us to understand the distribution of collaborating authors and the potential research topics defined by the collection of papers.

1) Analysis: We compute $diversity$ distribution for each cluster in each time snapshot. The results are presented in Figure 3 and Figure 4. Figure 3 is the mapping from authors network to papers network, while Figure 4 is the mapping from papers network to authors network. Due to the restriction of space, we only present results from 1980-1999. Besides the $diversity$ distribution of all clusters in the network, we also plot the influential nodes clusters $diversity$ distribution. We compute authors' influence based on PageRank [12] and H-Index [13]. The h-index attempts to measure both the productivity and impact of the published work of a scientist or scholar. The index is based on the set of the scientist's most cited papers and the number of citations that they have received in other publications. The index can also be applied to the productivity and impact of a group of scientists, such as a department or university or country. PageRank works as a topological measurement for the importance of an author in the collaboration networks, while h-index acts as an external metric for measuring the prominence of an author based on their publication citations.

Two different types of measurements are introduced to minimize the bias/noise introduced by prominence measurement methods, which makes sure that our conclusion and observation are solid and prominence measurement does not distort our experiments.

In the process of network evolution, the $diversity$ distribution maintains the similar properties, which can be drawn trivially from the shapes of distribution, this holds for both authors collaboration networks and papers co-author networks. Another interesting observations is, for influential authors (either identified by PageRank or H-index), they tend to have larger $diversity$ values than common authors; we can see majority of them have large $diversity$ values. While this situation is not significant for influential papers, they do not

Fig. 3. Authors Clusters Diversity Distribution

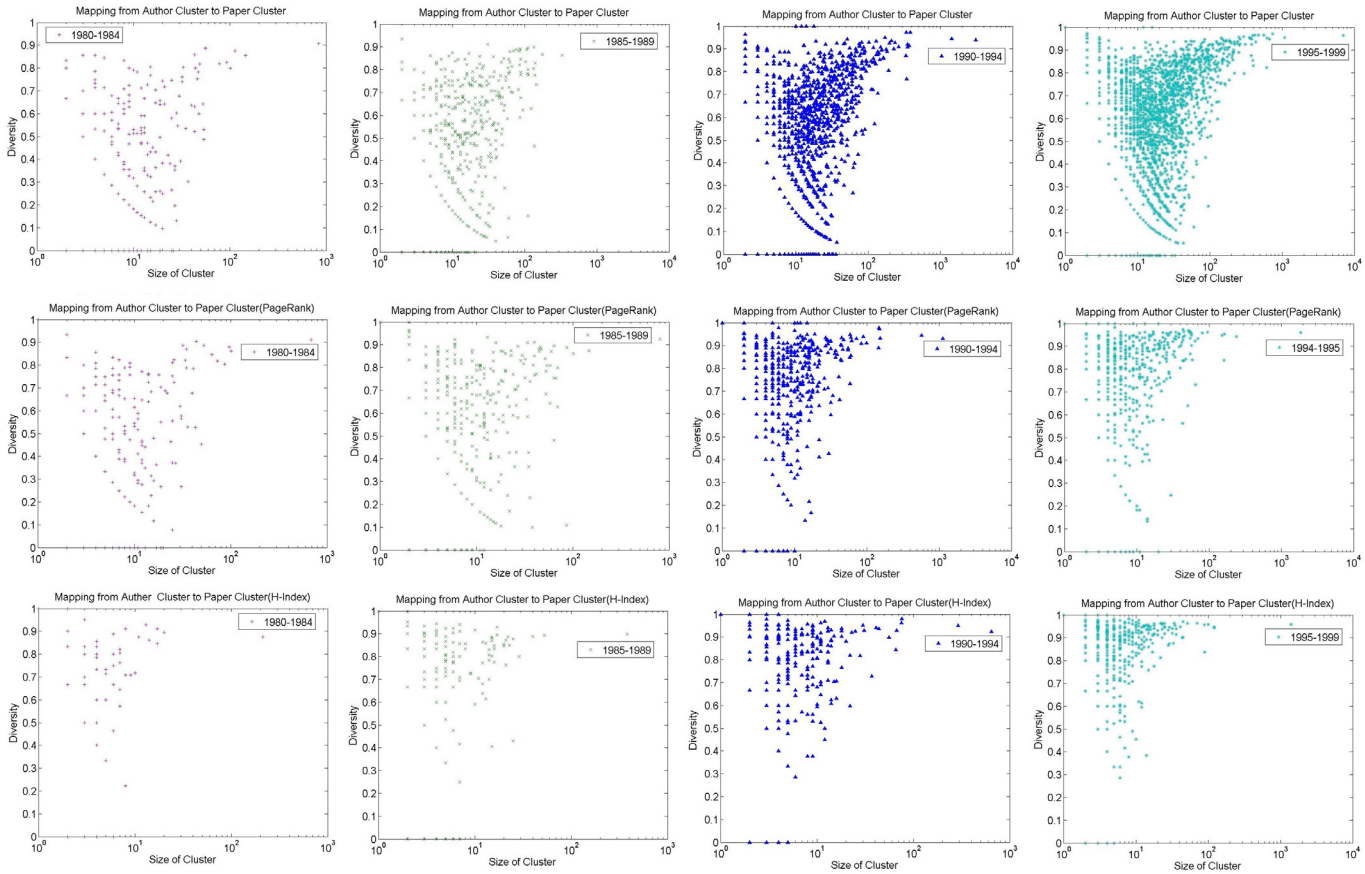


Fig. 4. Papers Clusters Diversity Distribution

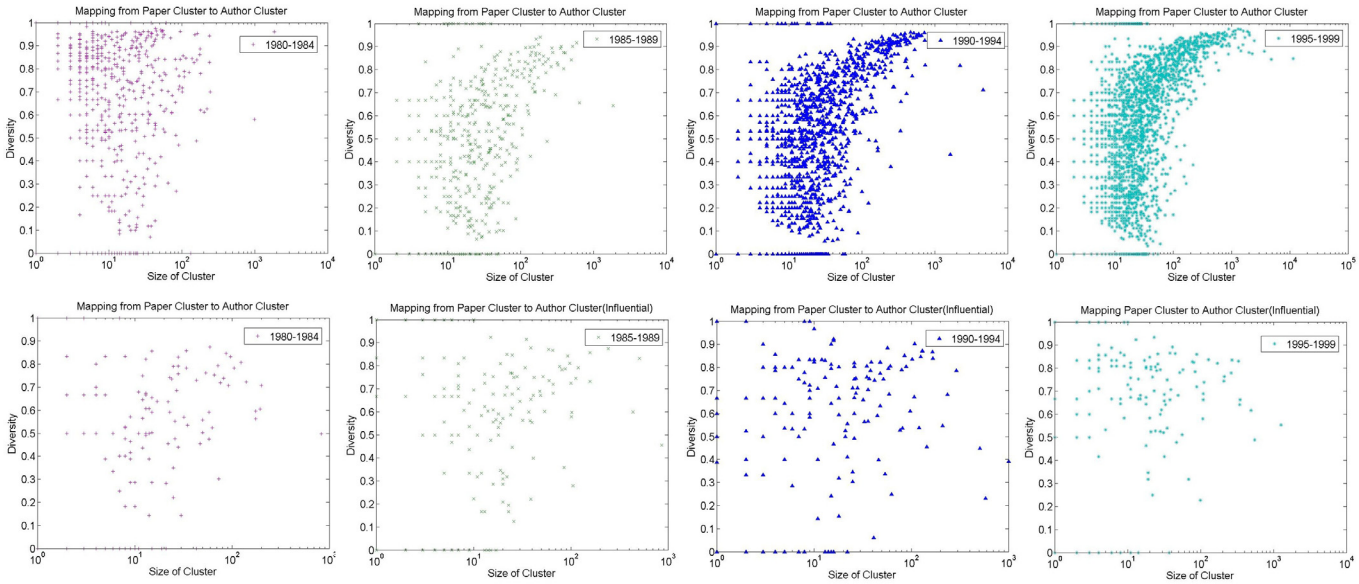
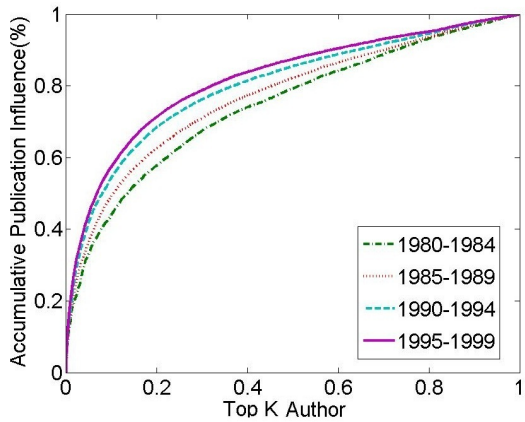


Fig. 5. Top K Percentage Author's Publication Influence



follow the same distribution as that of influential authors. However, there is still a trend that influential papers' $diversity$ values become larger and larger over time.

In our observation, the diversity distribution does not necessarily imply a larger value for a larger authors' cluster. However, as one would expect larger paper clusters' have higher probability to have larger diversity values. With influential authors' clusters diversity distribution, we can see that influential authors tend to have larger diversity values, which means influential authors incline to publish papers/artifacts in large number of research areas while general/common authors tend to focus on limited number of research areas. As for the paper diversity distribution, this phenomenon is not significant, we can conjecture that influential authors have influence over different research areas, while influential publications come from the efforts of authors in closely related research areas. In conclusion with the increasing influence of authors, especially for prominent authors, they tend to work in a more interdisciplinary circumstance, while an influential paper generally comes from efforts of experts in similar research areas.

B. Authors Prominence and Communities Evolution

In order to verify the behavior of influential authors in the evolution of networks, we look at the top k percent authors' publication influence over time. The results are presented in Figure 5. In observation we can see a large group of influential authors have influence in the process of network evolution. The rankings of authors or papers are based on the results of PageRank algorithms. In Figure 5, the top k percent authors are identified based on their PageRank scores. As we know there is link/mapping between authors and papers based on author-paper relationship, thus for the top k percent authors there is a set of corresponding papers, and we can sum up these papers' PageRank scores in paper co-author network as their cumulative publication influence. This measurement reveals the importance of authors in terms of their publications, and also confirms the conclusion of Guo et al. [14].

Guo et al. suggested that a small set of power actors in the network can dominate when the network fitting a power-law (DBLP network degree distribution follows power-law distribution). This also suggests that this small set of power actors in networks have influence on the evolution of communities in the network, however to prove this is difficult. To show that there is mutual effect between the influential authors development and the evolution of their corresponding communities, we conduct several experiments to identify the influence of important authors on the evolution of communities over time.

We slice the authors' collaboration network into 5 years snapshot networks, denote as G_1, G_2, ..., G_K; for each snapshot networks we run PageRank algorithm and Walktrap algorithm, and therefore for each time snapshot we have author's PageRank scores and their corresponding communities PageRank scores (simply the sum of PageRank scores of authors in the community). For each node (author) v we have an individual influence vector and a community influence vector.

- $influence(v) = (score_{G_1}(v), score_{G_2}(v), ...)$
- $influence(C_v) = (score_{G_1}(C_v), score_{G_2}(C_v), ...)$

Each element in the individual influence vector states the *PageRank* score of the node v in corresponding snapshot network, while each element in the community influence vector is the sum of *PageRank* scores of authors within the community.

By computing the pearson correlation of these two time series data we can see how an individual's development is related with its community's evolution over time.

For Figure 6 we can see that the influence of influential authors (top K by PageRank score) are highly correlated with the evolution of clusters where they belong to, the pearson correlation scores are mostly between 0.25-0.9, median value is 0.55; while for the general/common authors, these pearson correaltion coefficient scores are mostly between 0 and 0.18 (any values larger than 0.2 are outliers in the boxplot), which means these non-influential individuals neither have high impact on the evolution of clusters over time nor have been influenced significantly by the evolution of clusters over time (Figure 7). In constrast from Figure 6 we can conclude these influential/prominent authors not only influence the community evolution but also being influenced significantly by community in the process of evolution.

VI. Prominent Papers and Academic Research Trend

In this section, we introduce a method to measure the importance of an artifact or a paper in DBLP bibliographic network. There are two ways to measure the prominence of an artifact or paper, 1) measure paper's importance in paper co-author networks and use their PageRank scores as prominence value; 2) measure paper's importance in paper citation network. Both methods have obstacles, for paper co-author networks, papers will have high PageRank scores if they are written by many authors, while for citation network the problem resides in the completeness of data, in current DBLP dataset most citation information are lost. In this section

Fig. 6. Influential Authors Evolution and Corresponding Clusters Evolution

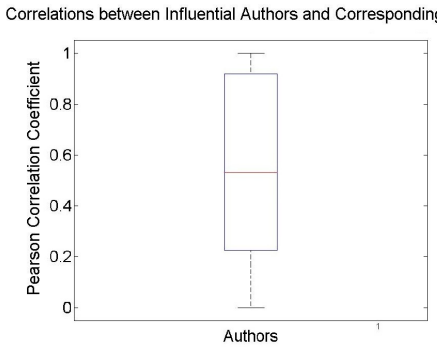
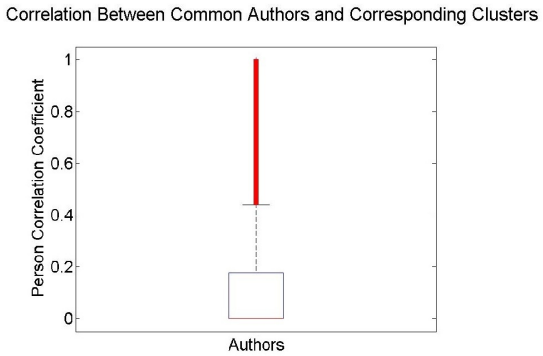


Fig. 7. Common Authors Evolution and Corresponding Clusters Evolution



we employ the incomplete citation information to measure the prominence of paper, and the results are promising. We also show that the small set of influential papers can reveal the trend of research topics over time precisely.

A. Diffusion

We employ breadth-first search procedure to propagate the influence of a paper that a restricted random walk starting at v ends at u in l steps or fewer using link weights as transition probabilities, which is described in Algorithm 1. By adding up all the influence scores propagated from a paper v to all other papers, we can get the total influences of paper v over the whole network, which can be used for paper prominence comparison.

B. Experiment Set-up

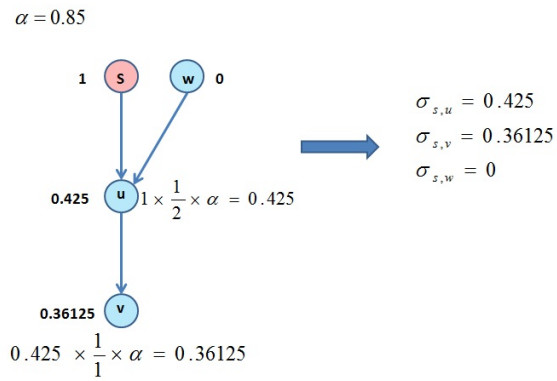
For the paper citation network from 1974 to 2009, we split it using 5 years time span. Each snapshot network only represents the citation activities happening in 5 years time span. This is intuitive — if the influential paper of one specified research area was not cited for a long time, this means the research area is not so popular in recent years. In this way we can capture the trend of research topic by papers' influence over time.

From the matrix M described in the last subsection, we can compute the total influence for each paper, denoted as $sum(\sigma_v)$ for a paper v. For each snapshot network, we can extract a set of papers, which have top K total influence

Algorithm 1 Diffusion

Input: $G \leftarrow$ input graph. $G = (V, E)$.
Initialize: $\alpha \leftarrow$ decay factor, traditionally set to 0.85.
Initialize: $\sigma_{ij} \leftarrow 0, \forall i, j \in V$.
for $\forall u \in V$ **do**
  $BFS(u) \leftarrow$ BFS tree from i with levels P_0 to P_h. Note, $P_0 = \{u\}$.
  If $v \in \Gamma(u)$, set $\sigma_{uv} \leftarrow \frac{\alpha}{d(v)}$.
  $\forall w \in \Gamma(v)$ and $\forall l$, $l = 0, 1, \ldots, h$ such that $\forall v \in P_l$ and $w \in P_{l+1}$, set $\sigma_{uw} \leftarrow \max_v [\frac{\alpha}{d(w)} \cdot wt(v, w) \cdot \sigma_{uv}]$.
end for
Output: $\mathbf{M} \leftarrow [\sigma_{ij}]$

Fig. 8. Diffusion Toy Example



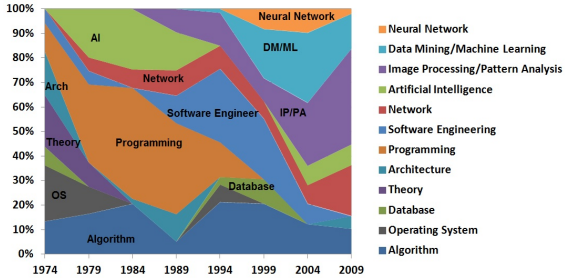
$sum(\sigma)$, denote as $\boldsymbol{S}$ (prominent paper set); while for non top K papers, we classify them into another set, called $\boldsymbol{T}$ (non prominent paper set). As we know in matrix M, there is influence score $\sigma_{u,v}$ from node u to node v. Therefore by using these information, we can construct a bipartite network between sets $\boldsymbol{S}$ and $\boldsymbol{T}$. By employing RankClus [15] algorithm we can then classify the top K papers into several clusters.

C. Research Trend Analysis

In this section we introduce a method to extract research topics trend of computer science from DBLP paper citation networks based on the experimental configuration discussed above. By employing the RankClus [15] algorithm we can classfiy the top K papers into several clusters, and using terms and conferences information of these papers we label a topic for each cluster identified. Considering the total size of prominent papers for each snapshot is K, and by counting the cluster size for each topic (label) we can generate a stacked area chart over time (Figure 9).

From the stacked area chart in Figure 9 the trends of important research areas of computer science are identified precisely. Take Programming for example, the arise of Programming research is initiated by the introduction of C language, Pascal and SQL developed around 1970s; and then with the work of important languages, such as C++ and Ada, Programming

Fig. 9. Research Trend Analysis



research was at the top of its history; when in 1990s by the population of web and introduction of Java, Programming research still played an important role in computer science. While after 2000 it is not so popular in academic area as before. This trend of Programming Research shown in Figure 9 precisely follows the history described in Wikipedia. Another example is the Software Engineering, which was popular between 1990-2000, after that the popularity damped. With careful analysis and comparison with real world research trend of computer science, we can find that the trends identified by our method are promising for all the research areas listed.

VII. Community Based Evolution

In this section, we analyze evolution of a network as a function of the communities of objects (papers). We divide the DBLP network into partially overlapping time slices, each consecutive time slice has a 5 year overlap. This allows us to follow the evolution of papers in a specific area. We will denote these networks by $\mathcal{D}^1, \ldots, \mathcal{D}^n$. We compute the paper to paper graph for each network $\mathcal{D}^i$ and find the clusters of papers for that network. We will use the notation $C_1^i, \ldots, C_{m^i}^i$ for the clusters in network i.

We match two clusters C_p^i and C_q^{i+1} if they share some common papers. Based on this match, we track the evolution of a cluster in the following time slices. Our hypothesis is that most influential papers need a community to continue working on related topics and draw interest to that specific problem. Note that this may not happen as soon as the paper is published. As a result, this trend may not be visible in the publication record, but in the citation record. We investigate the cases where the publication record shows this trend. Figure 10 shows an example evolution from a cluster containing two influential papers to other influential papers.

We compute a score for each cluster based on the number of clusters that have large and sustained overlap over the different time slice. The community diffusion score, CDS is computed as follows:

$$\begin{aligned} CDS(C_p^n) &= 0 \\ CDS(C_p^i) &= \frac{1}{m^i} \sum_{C_q^{i+1} \cap C_p^i \neq \emptyset} (1 + CDS(C_q^{i+1})) \end{aligned}$$

This measure favors clusters that overlap with a large number of clusters in each different time slice. The longer the time period, the higher the CDS score. To evaluate the effectiveness of the CDS score, we extract the top 10,000 most influential papers from citeseer [1] and exclude all papers from the last time slice, for which our CDS score would be 0. Of the remaining, 3,590 matched with our dataset. In Figure 11, we plot each cluster containing an influential paper. The x-axis shows the CDS score of the cluster, and y-axis shows the number of influential papers in that cluster. For all clusters containing more than one influential paper, the CDS value is a good indicator of its importance. One possible explanation is that the largest clusters are the ones containing the influential papers. To check this, we plot the number of influential papers as a function of cluster size in Figure 12. We see that while largest clusters have a strong correlation, there is no correlation for small clusters. And the majority of the clusters are small in size.

Thus, this experiment shows that while the influential individual or high impact paper is important, the community attachment and growth also has an impact on the individual's prominence. Prominence is indeed a co-evolving process, and an individual cannot do without the community, and if the community does not rally behind a paper or a topic championed by an individual, prominence might be affected.

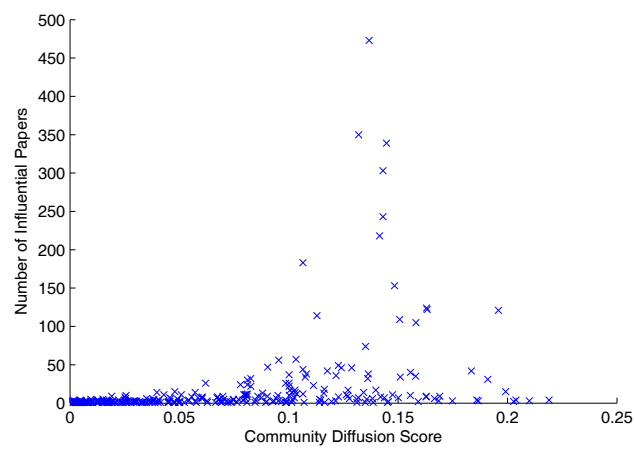


Fig. 11. The community diffusion score of all the clusters containing an influential paper vs the number of influential papers in that cluster.

VIII. Conclusions

This paper focused on the fundamental questions of: how people and their artifacts gain prominence in collaborative networks? We posited that prominence is a co-evolving process — that is the individuals and their papers affect their position in the network, and the communities that are crated around collaboration or research topic also in return affect the individual's or paper's prominence. We proposed a metric, *diversity*, that captures the distribution of research interests and topics of authors. We find that prominent people have more diverse interests of research topics, and the influential papers

[1] citeseer.ist.psu.edu/stats/citations

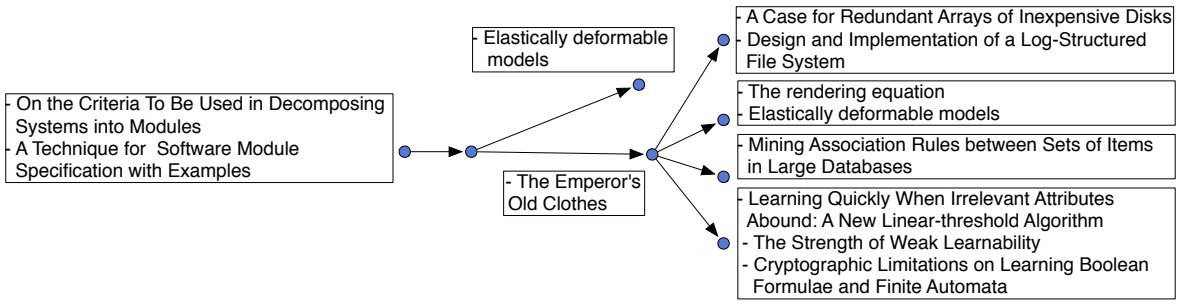


Fig. 10. An example evolution of research areas.

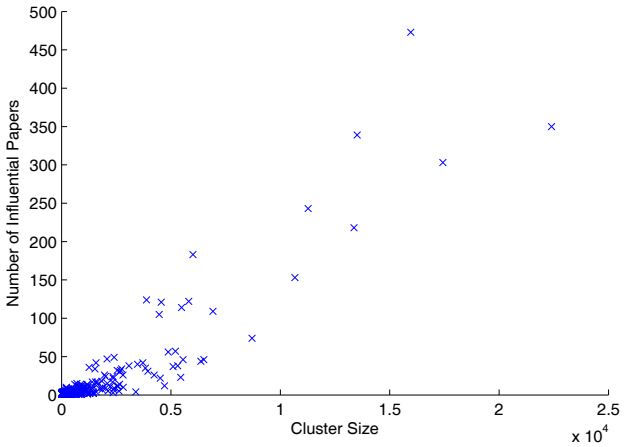


Fig. 12. The number of influential papers as a function of the size of the cluster. Note that there is a correlation with size only for the largest clusters.

generally come from efforts of people from similar research area. Additionally, by looking at the correlation between individuals' evolution and their communities evolution, we conclude that prominent people have much more significant correlation with their communities than non-prominent people. With a well designed method in Section VI we precisely identify the trend of Computer Science researches, which also proves that prominence of papers are heavily impacted by the research trend over time. We also demonstrate that clusters or communities have an effect on the prominence of an individual or paper, establishing the co-evolutionary forces propelling prominence in a network.

ACKNOWLEDGMENTS

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

REFERENCES

[1] P. Pons and M. Latapy, "Computing Communities in Large Networks Using Random Walks Computer and Information Sciences - ISCIS 2005," in *Computer and Information Sciences - ISCIS 2005* (p. Yolum, T. Güngör, F. Gürgen, and C. Özturan, eds.), vol. 3733 of *Lecture Notes in Computer Science*, ch. 31, pp. 284–293, Berlin, Heidelberg: Springer Berlin / Heidelberg, 2005.

[2] J. Paul, "Distribution de la flore alpine dans le bassin des dranses et dans quelques rgions voisines," Sciences Naturelles 1901, pp. 241–272, 1901.

[3] S. Adalı, X. Lu, M. Magdon-Ismail, and J. Purnell, "Prominence ranking in graphs with community structure," in *International AAAI Conference on Weblogs and Social Media (ICWSM'11)*, 2011.

[4] Kossinets and D. Watts, "Empirical analysis of an evolving social network," *Science*, vol. 311, no. 5757, pp. 88–90, 2006.

[5] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: densification laws, shrinking diameters and possible explanations," in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, KDD '05, (New York, NY, USA), pp. 177–187, ACM, 2005.

[6] 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*, KDD '06, (New York, NY, USA), pp. 44–54, ACM, 2006.

[7] D. Chakrabarti, R. Kumar, and A. Tomkins, "Evolutionary clustering," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '06, (New York, NY, USA), pp. 554–560, ACM, 2006.

[8] L. Tang, H. Liu, J. Zhang, and Z. Nazeri, "Community evolution in dynamic multi-mode networks," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08, (New York, NY, USA), pp. 677–685, ACM, 2008.

[9] M. L. N. Xiaoming Liu, Johan Bollen and H. V. de Sompel, "Co-authorship networks in the digital library research community," *Information Processing & Management*, vol. 41, pp. 1462–1480, 2005.

[10] Y. Sun, J. Han, P. Zhao, Z. Yin, H. Cheng, and T. Wu, "Rankclus: integrating clustering with ranking for heterogeneous information network analysis," in *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, pp. 565–576, 2009.

[11] Y. Sun, Y. Yu, and J. Han, "Ranking-based clustering of heterogeneous information networks with star network schema," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '09, pp. 797–806, ACM, 2009.

[12] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web.," Technical Report 1999-66, Stanford InfoLab, November 1999. Previous number = SIDL-WP-1999-0120.

[13] J. E. Hirsch, "An index to quantify an individual's scientific research output," PNAS 2005, p. 165691657, 2005.

[14] L. Guo, E. Tan, S. Chen, X. Zhang, and Y. E. Zhao, "Analyzing patterns of user content generation in online social networks," KDD 2009, (New York, NY, USA), pp. 369–378, ACM, 2009.

[15] Y. Sun, J. Han, P. Zhao, Z. Yin, H. Cheng, and T. Wu, "Rankclus: integrating clustering with ranking for heterogeneous information network analysis," EDBT '09, 2009.