

Evaluating cooperation in communities with the k -core structure

Christos Giatsidis

LIX

École Polytechnique

Palaiseau Cedex, France

Email: xristosakamad@gmail.com

Dimitrios M. Thilikos

Department of Mathematics

National & Kapodistrian University of Athens

Athens, Greece

Email: sedthilk@math.uoa.gr

Michalis Vazirgiannis

Department of Informatics,

Athens Univ. of Economics

Athens, Greece

Email: mvazirg@aueb.gr

Abstract—Community subgraphs are characterized by dense connections or interactions among its nodes. Community detection and evaluation is an important task in graph mining. A variety of measures have been proposed to evaluate the quality of such communities. In this paper, we evaluate communities based on the k -core concept, as means of evaluating their collaborative nature - a property not captured by the single node metrics or by the established community evaluation metrics. Based on the k -core, which essentially measures the robustness of a community under degeneracy, we extend it to weighted graphs, devising a novel concept of k -cores on weighted graphs. We applied the k -core approach on large real world graphs – such as DBLP and report interesting results.

Index Terms—Community evaluation, k -core, graph mining, co-authorship graphs

I. INTRODUCTION

Large and evolving graphs constitute an important element in current large scale information systems. Common cases of such graphs are the Web graph, social networks, citation graphs, CDRs (call data records) where nodes (featured with attributes - in some cases with a large cardinality) are connected to each other with directed edges representing a relation such as endorsement/recommendation/friendship. In all cases and to a great extent due to the economic aspects of these networks the ranking of individual nodes is a cornerstone concept.

Graphs of real data with community structure have a broad distribution on the degree a node may display. As it is pointed out in [2], nodes of low degree coexist with nodes of high degree making the graph in-homogeneous both globally and locally which usually indicates order and structural behavior i.e. communities.

Community sub-graphs are characterized by dense connections or interactions among its nodes. Community detection and evaluation is an important task in graph mining. A variety of measures have been proposed to evaluate the quality of such communities. In this paper, we evaluate communities based on the k -core concept, as means of evaluating their collaborative nature - a property not captured by the single node metrics or by the established community evaluation metrics.

Our contributions lie in the following:

- A novel metric for evaluating the cohesiveness of communities based on the k -core structure.

- An innovative extension of the k -core concept assigning weights on the edges. The weights represent the degree of cooperation among the two connected vertices.
- We develop an extended experimental evaluation in the case of the DBLP co-authorship graph that results in very interesting findings.

In the rest of this paper we:

- Provide motivation for the identification of the most cohesive part of a community graph and give special emphasis to the DBLP coauthor ship graph.
- Introduce step by step the fractional core method by firstly presenting the k -core algorithm and its complexity on un-weighted graphs and then expanding the graph model by adding weights and altering the basic algorithm accordingly.
- Present and describe the DBLP dataset and the application of the k -core algorithm on its coauthor ship graph. Following we filter the DBLP dataset, to exclude abnormalities on the general behavior of the graph, and apply again the k -cores. Lastly the proposed fractional core method is utilized and a comparison of the 3 different versions of results is made.
- Finally we present our conclusion on this novel application of k -cores and its variation for the evaluation of communities and its members.

II. RELATED WORK

A thorough review on community detection in graphs is offered by Fortunato in [6]. In that work techniques, methods and data sets are presented for detecting communities in sociology, biology and computer science, disciplines where systems are often represented as graphs. Most existing relevant methods are presented, with a special focus on statistical physics, including discussion of crucial issues like the significance of clustering and how methods should be tested and compared against each other.

The k -cores are fundamental structures in graph theory and their study dates back to the 60's [4], [14], [8]. The existence of k -cores of large size in sufficiently dense graphs has been theoretically studied by [12] for random graphs generated by the Erdős-Rényi model [5]. As shown in [12], a k -core whose size is proportionate to the size of G (i.e. a “giant” k -core)

appears in a random graph with n vertices and m edges when m reaches a threshold $c_k \cdot n$, for some constant c_k that depends exclusively on k .

Studying the general behavior and properties of real graphs, both weighted and unweighted, is the subject of [9] where a pattern on the behavior of connected components over time is observed and a upon that is build a generative model.

In recent literature various metrics are proposed relevant to the graph structure of a social network. Such are ‘‘Betweenness’’ [15], ‘‘Centrality’’ [11], ‘‘Clustering coefficient (a measure of the likelihood that two associates of a node are associates themselves. A higher clustering coefficient indicates a greater ‘‘cliquishness’’, i.e. cohesion degree or density. Of special interest here is the eigenvector centrality- a measure of the importance of a node in a network. It assigns relative scores to all nodes in the network based on the principle that connections to nodes having a high score contribute more to the score of the node in question. Other measures include ‘‘path length’’ (i.e. distances between pairs of nodes in the network), ‘‘prestige/authority’’, a measure in directed graphs to describe a node’s centrality and ‘‘radiality’’, a notion representing an individual’s network to reach out into the network and thus is influential. Other interesting measures include ‘‘Structural cohesion’’ - the minimum number of members who, if removed from a group, would disconnect the group [10]. In [7] an idea similar to the k -cores is used to filter out less significant nodes, by pruning them out. The main difference to our approach is that it removes only a sufficient portion of the nodes. The cores are then fed to a generalized HITS algorithm used to expand the communities within them. In [3] greedy approximation algorithms are proposed for finding the dense components of a graph. Both undirected and directed graphs are examined. In the case of directed graphs the vertices are divided in hubs (S) and authorities (T), then based on a value of $|S|/|T|$ a greedy algorithm removes the vertex of minimum degree from either S or T until both sets are empty. Also in [13] the subject of finding dense subgraphs based on query nodes is studied where the subject is to find a community that contains certain given nodes.

III. THE FRACTIONAL CORE METHOD

Motivation for finding the most cohesive part of community graph – special emphasis to DBLP co-authorship graph

A. k -core as cohesion measure for cohesive co-authorship communities

Let G be a simple undirected graph. We denote by $\Delta(G)$ the minimum degree of a vertex in G . The degeneracy of G is defined as follows.

$$\delta^*(G) = \max\{\Delta(H) \mid H \subseteq G\}$$

Also we define the k -core of a graph as the maximum size sub graph H of G where $\Delta(H) \geq k$, where k is a positive integer. It can be proved that such a sub graph is unique.

DEFINITION 1 (Vertex core number) *The core number of a vertex v of G is the maximum k for which v belongs in the*

k -core of G .

DEFINITION 2 (Subgraph core number) *The core number of a subset S of the vertices of G is the maximum k for which all vertices of S belong in the k -core of G .*

The algorithm for computing the k -cores follows:

Procedure $Trim_k(D)$

Input: An undirected graph D and positive integer k

Output: k -core(D)

1. let $F \leftarrow D$.
2. **while** there is a node x in F such that $\deg_F(x) < k$
 delete node x from F .
3. **return** F .

The k -core algorithm is of low complexity thus computations are feasible even in large scale digraphs.

We consider the DBLP bipartite graph $G_{\text{DBLP}} = (A, P, E)$ where A is the set of authors, P is the set of papers, and E is a set of edges. Each edge $\{x, y\}$ (where $x \in A$ and $y \in P$) expresses the fact that x is one of the authors of paper y . We also assume that all the papers are written by at least two authors, i.e., vertices in P have degree at least 2.

Our first wave of experiments considers the graph $G_{\text{DBLP}} = (A, P, E)$ and build the co-authorship graph

$$H_{\text{DBLP}} = (A, \{\{x, x'\} \mid \exists y \in P : \{x, y\}, \{x', y\} \in E\}),$$

i.e., two authors are made adjacent if they appear as co-authors in some paper.

We consider $\delta^*(H_{\text{DBLP}})$, the core number of each vertex/set of vertices in H_{DBLP} , and the core of $\delta^*(H_{\text{DBLP}})$ in order to evaluate the collaboration behaviors in the DBLP graph.

However, this is not an entirely satisfactory evaluation as papers of many co-authors have equal weight in this measuring as papers with fewer. For this reasons we introduce below a more refined way to define cores starting from bipartite graphs (such as the DBLP graph).

B. Fractional k -core for weighted graphs

Let (A, P, E) be a bipartite graph where all vertices in P have minimum degree 2. If, we define the neighborhood $N(x)$ of x as the set of $y \in P$ for which $\{x, y\} \in E$. Symmetrically, we define the neighborhood $N(y)$ of a paper $y \in P$. Also, given an author x we denote by $E(x)$, the set of all edges with x as an endpoint.

DEFINITION 3 (Co-authorship edge weight) *We define the weighted co-authorship graph by taking H_{DBLP} , as defined before, as setting up a rational weighing function $\mathbf{w} : E \rightarrow \mathbb{Q}$ on the edges of H_{DBLP} as follows. For every edge $e = \{x, x'\}$ we set*

$$\mathbf{w}(e) = \sum_{y \in N(x) \cap N(x')} \frac{1}{N(y)}$$

Notice that, $\sum_{e \in H_{\text{DBLP}}} \mathbf{w}(e) = |V(P)|$, i.e. the sum of all the weights on the edges is the size of the graph in terms of vertices.

Let (G, w) be a graph together a weight function w assigning rational numbers on its edges.

DEFINITION 4 (*Vertex fractional degree*) We define the fractional-degree of x in (G, w) as

$$\text{deg}_{G, \mathbf{w}}(x) = \sum_{e \in E(x)} \mathbf{w}(e)$$

If (H, \mathbf{w}') is a sub graph of (G, w) where \mathbf{w}' is the restriction of \mathbf{w} to the edges of H , we define $\Delta(H, \mathbf{w}') = \min\{\text{deg}_{H, \mathbf{w}'}(x) \mid x \in V(H)\}$.

Definition 5. Graph fractional-degeneracy: Consider now the graph H_{DBLP} together with its weight function w . We define its fractional-degeneracy as follows:

$$\delta^*(G, \mathbf{w}) = \max\{\Delta(H, \mathbf{w}') \mid (H, \mathbf{w}') \text{ is a subgraph of } (G, \mathbf{w})\}.$$

Let $k \in \mathbb{Q}$. Then the k -core of (G, w) is the maximal sub graph (H, \mathbf{w}') of (G, \mathbf{w}) where $\Delta(H, \mathbf{w}') \geq k$.

The *core number* of a subset S of the vertices of an edge-weighted graph (G, \mathbf{w}) is the maximum rational number k for which all vertices of S belong in the k -core of G . The core number of a vertex x is defined as before, by taking $S = \{x\}$.

Fractional cores are essentially defined for hyper-graphs where hyper-edges express the groups of distinct elements of a set (in our case, papers written by authors). For simplicity, we study such hyper-graphs using their incident (bipartite) graphs. This directly produces weighted graphs where fractional cores are detected. Essentially, our work is about fractional cores on weighted graphs where the weights emerge by hyper-graphs (and their incident bipartite graphs).

IV. EXPERIMENTAL EVALUATION ON DBLP

In this section we present the evaluation of the above defined framework on the DBLP coauthor ship graph H_{DBLP} . We compute the k -core and fractional k -core on different versions of H_{DBLP} and thus compute the relevant sub graphs finding thus in each case the most dense communities in terms of co authorship collaboration.

A. Data set description and preprocessing

The DBLP dataset is freely available in XML format at <http://dblp.uni-trier.de/xml/>. From this dataset we extract the un-weighted and weighted undirected graph by considering a coauthor ship of a paper from two or more authors as an edge that connects each pair of the co-authors from the paper. In the current snapshot there were approximately 825K authors.

In Figure 1 the reader can see the distribution of the number of authors per paper in the DBLP graph. It is clear the vast majority of the papers are authored by few authors. There are some extremities where one specific paper has 114 coauthors!

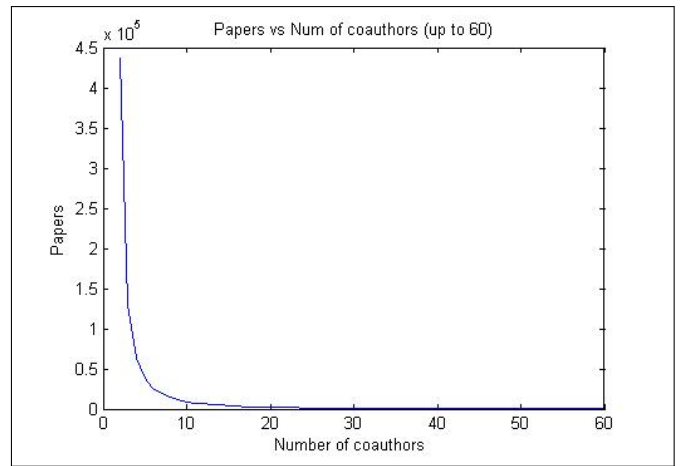


Figure 1. Coauthor ship distribution per paper.

B. k -cores on unfiltered graph

We apply of the k -core algorithm on the graph considering all the papers regardless of the number of authors (unfiltered). In Figure 2. we can see the distribution of the sizes of the k -cores in the unfiltered graph case.

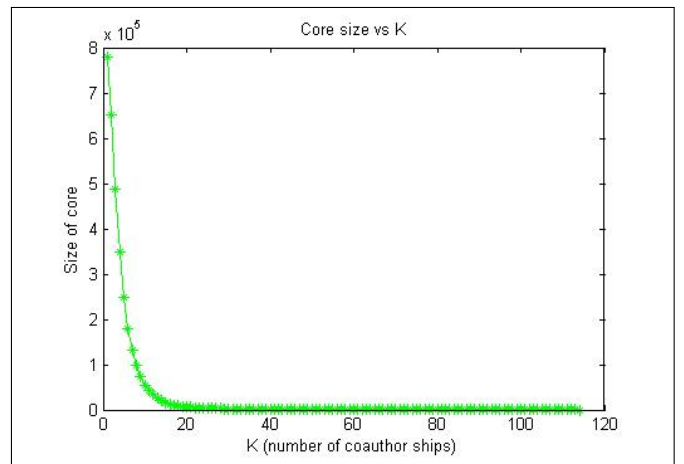


Figure 2. Distribution of the k -core sizes in the unfiltered DBLP coauthorship graph.

Here we will elaborate on the semantics of a k -core in the co-authorship graph. As mentioned previously one paper with a large number of coauthors can “push” authors with otherwise low coauthor ship (or even practically none) to the densest k -core. For example at $k=113$ we have 114 authors that all of them have participated in the same publication and some of them do not appear anywhere else on the dataset.

Following we present a ranking of a few selected authors based on the maximum core they belong to (Table I).

The results of the k -core application on the *unfiltered* graph apparently are extremely biased, i.e. a 113-core in the graph with all authors of a single paper, regardless of their other publication activity.

| Name of author | k -core |
|---------------------------|-----------|
| Serge Abiteboul | 28 |
| Christos Faloutsos | 28 |
| Gerhard Weikum | 22 |
| Christos H. Papadimitriou | 17 |
| Paul Erdős | 16 |

Table I
 k -CORE RANKING UNFILTERED

This motivates us to consider filtering out papers with extremely high number of coauthors. In this case the graph is formed by the authors of the papers whose number of co-authors is within the 99% of the distribution shown in Figure 1. This leaves out papers with more than 15 coauthors – i.e. less than 0.01% of the papers in DBLP bear more than 15 coauthors each. We call this version of the graph *filtered* further on.

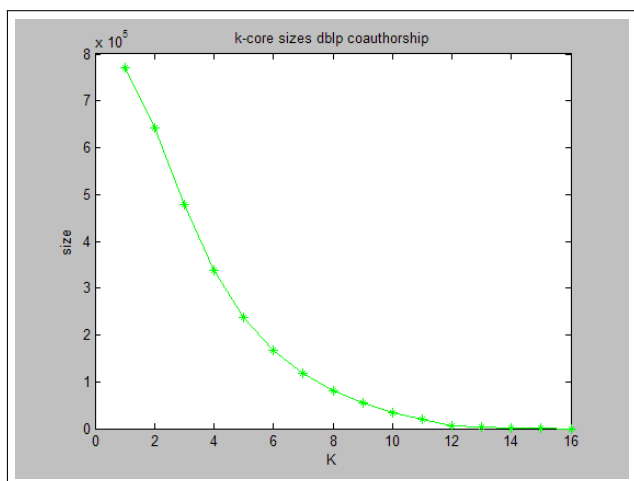


Figure 3. Distribution of the k -core sizes in the filtered DBLP coauthorship graph.

We applied the k -core algorithm on the filtered graph – considering papers with max 15 coauthors. The distribution of the resulting k -core sizes appears in Figure 3. In the filtered data the most dense core is at $k=15$, including size of 76 authors. These authors appear in TABLE II.

By excluding papers we, expectedly, get many of the authors “move down” in cores of smaller k values. In TABLE III. we can see the new rankings of the same authors. As we can see the top authors get accumulated in the second most cohesive core. This means that, even though these authors are very collaborative, some of their coauthors had only a few collaborations on papers with a large number attributed authors. The most cohesive k -core (15) has authors that are highly collaborative with other authors that behave in a similar way. It is interesting that for some authors i.e. Tanenbaum the vertex degree in the filtered case is much lower (12) that in the unfiltered one (48) apparently due to his participation in multi-author papers that were filtered out.

| | | |
|-----------------------|-----------------------|----------------------|
| Kurt Mehlhorn | Micha Sharir | Pankaj K. Agarwal |
| Rolf Klein Mark | H. Overmars | Herbert Edelsbrunner |
| Stefanie Wührer | Jack Snoeyink | Joseph O'Rourke |
| Subhash Suri | Otfried Cheong | Hazel Everett |
| Sylvain Lazard | Helmut Alt | Emo Welzl |
| Günter Rote | Leonidas J. Guibas | Chee-Keng Yap |
| Danny Krizanc | Pat Morin | Jorge Urrutia |
| Diane L. Souvaine | Ileana Streinu | Dan Halperin |
| Hervé Brönnimann | Joseph S. B. Mitchell | David Eppstein |
| Erik D. Demaine | Olivier Devillers | Sándor P. Fekete |
| Henk Meijer | Sariel Har-Peled | John Hershberger |
| Alon Efrat | Stefan Langerman | Bernard Chazelle |
| Joachim Gudmundsson | Giuseppe Liotta | Sue Whitesides |
| Christian Knauer | Raimund Seidel | Michiel H. M. Smid |
| Tetsuo Asano | David Rappaport | Vera Sacristan |
| Hee-Kap Ahn | Prosenjit Bose | Michael A. Soss |
| Godfried T. Toussaint | Marc J. van Kreveld | Martin L. Demaine |
| Ferran Hurtado | Timothy M. Chan | Oswin Aichholzer |
| Bettina Speckmann | Jeff Erickson | Therese C. Biedl |
| Greg Aloupis | David Bremner | Anna Lubiw |
| Esther M. Arkin | Boris Aronov | Vida Dujmovic |
| Suneeta Ramaswami | Thomas C. Shermer | David R. Wood |
| Perouz Taslakian | John Iacono | Sergio Cabello |
| Sébastien Collette | Belén Palop | Mirela Damian |
| Jiri Matousek | Otfried Schwarzkopf | Richard Pollack |
| Mark de Berg | | |

Table II
15-CORE AUTHOR LIST FILTERED

| Name of author | k -core |
|---------------------------|-----------|
| Serge Abiteboul | 14 |
| Christos Faloutsos | 14 |
| Gerhard Weikum | 14 |
| Christos H. Papadimitriou | 14 |
| Paul Erdős | 14 |
| Tanenbaum | 12 |

Table III
 k -CORE RANKING FILTERED

C. Fractional cores on the weights graph

Here we articulate the need for assigning weights to the previously defined co-authorship graph. Assume two authors x, y that have coauthored some papers, therefore they are connected by an edge e . This co-authorship relation represents a strong collaboration among the two. This collaborative effort as it is evident is larger as the number of coauthored papers increases. On the other hand the effort to author a paper is naturally split among all papers coauthors (we assume in equal parts). This justifies the definitions in previous sections towards a weighted coauthor ship graph. Thus having defined the fractional degree of a vertex the k -core algorithm can easily be extended to compute degeneracies where k is a rational number. The new algorithm differs only in selecting an increment step for the rational limit and an extra computation - in the loop- for calculating the sum of the weights at each node.

In the case of indirect filtering that applies a low weight to papers with high number of coauthor ship. In Figure 4, we can see the distribution of the fractional k -cores sizes with k . The behavior is again exponential (in terms of decreasing k -core size with k). The maximum edge weight is 149.50 and the core contains one pair of authors (Sudhakar M. Reddy,

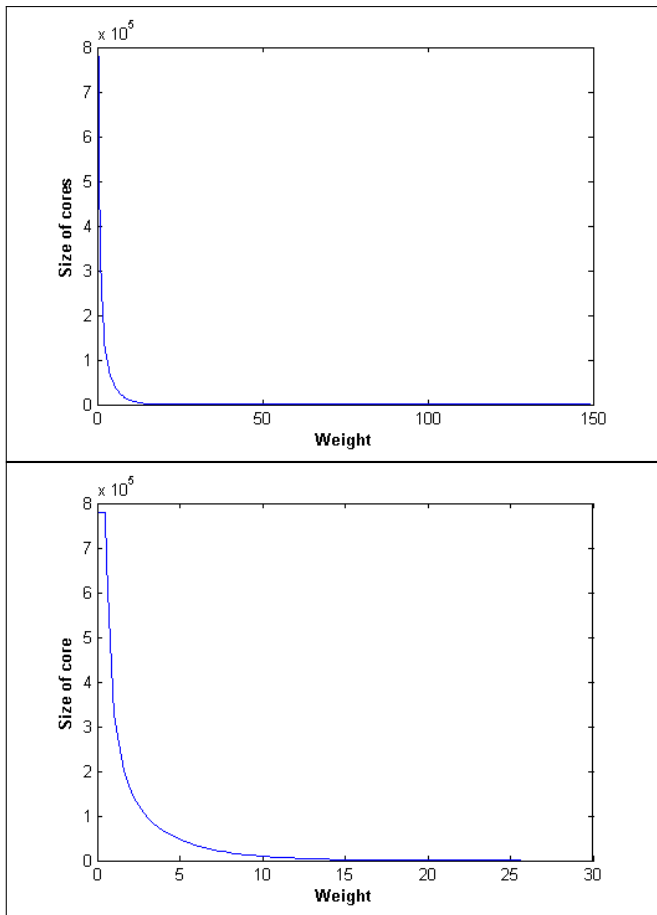


Figure 4. Distribution of the fractional k -core sizes in the DBLP coauthorship graph Up: full extent, Down: the less dense cores ($k < 26$).

Irith Pomeranz) whose publication record indeed verifies the claims as they have coauthored 373 papers, 256 of them as sole 2 coauthors! The next step where k -core size increases is $k = 77.79$ including the additional authors: Henri Prade, Didier Dubois whose strong collaboration is verified by the number of coauthored papers (223 according to DBLP) as well by the moderated number of coauthors in them.

In the upper part of Figure 4 we see the fully fledged distribution where at the bottom part we present the same distribution ignoring the extreme k values ($k > 26$) representing a very tiny part of nodes. and but the size of the cores has an immediate drop. (On the upper part we can see the full extent of the cores where in the lower one the weight axis has been cut off for better understanding of the “early” part of the behavior of the fractional cores). In this case the results differ to the previous unweighted cases. This is due to the weighting scheme. Indeed the assigned weight $w(e)$ is proportional to the number of papers they have coauthored and inversely proportional to the number of coauthors per coauthored paper. Thus $w(e)$ represents the “amount” of collaboration among x, y in terms of effort committed for common publications which is of course is normalized in each case by the number of contributing coauthors. This implies

that the best fractional k core communities contain authors that are intensively coauthoring with others but the number of coauthors is not high and thus the share of collaborative effort is high. Assuming an author x in H_{dblp} it should be stressed that her best hop-1 coauthorship k -core (i.e. immediate coauthors) are those that have at least k coauthors as well. This presents an interesting property that this structures evolves as a metric of collaboration where the collaborators should also be important with regards to this metric.

| Name of author | k -core |
|---------------------------|-----------|
| Christos H. Papadimitriou | 20.8 |
| Serge Abiteboul | 20.5 |
| Christos Faloutsos | 18.7 |
| Gerhard Weikum | 16.3 |
| Paul Erdős | 13.9 |

Table IV
 k -CORE RANKING FILTERED

In TABLE IV. are displayed the new rankings, of the previous sample of authors, based on the fractional cores computation. It is interesting that there is a different ranking in this case due to the weighting scheme that favors not only plethora of publications but also repetitive co-authorship with limited number of coauthors. In this case intensive collaboration with certain coauthors over a long series of publications increases the mutual edge weights and thus the ranking in the factorial k -cores.

In TABLE V we see the summary of the k -cores for the seminal author P. Erdős for all the graph versions (unfiltered, filtered, and factorial k -core). In the unfiltered case P. Erdős Vertex Core number is 16 (i.e. belongs to a 16-core) with 12802 more authors and his hop-1 community contains 20 names (coauthors) that as well belong to the same core (i.e. apparently P. Erdős has more coauthors whose Vertex Core number is less than 16 and therefore they do not survive until this level). In the filtered case P. Erdős Vertex Core number is 14 with 1236 more authors and his hop-1 community contains 20 names (coauthors) that as well belong to the same core. It is interesting that here the Vertex Core number is lower in comparison to the unfiltered case due to the removal of nodes that participate in papers with many other authors (>15). With regards to the factorial core structure the Vertex fractional degree is 13.9 with 2678 other coauthors. There is a large overlap in the hop-1 communities in all cases.

In Table VI we see the relevant data for fractional cores for a selection of well known and seminal authors representing their degree of collaborations with their coauthors. C. H. Papadimitriou has a top score in this measure (20.8) while having a very small but cohesive community of coauthors, with prominent example Yannakakis contributing an awesome weight (19.62) on the vertex fractional degree of Papadimitriou. This implies that they have coauthored many papers together (46) out of which more than 30 are coauthored by the two of them only! On the other hand G. Weikum has a much more distributed collaboration circle in terms of coauthors

| | max- k | Core size | hop-1 - Authors in core |
|---------------------|----------|-----------|--|
| unfiltered | 16 | 12802 | Boris Aronov, Daniel J. Kleitman János Pach, Leonard J. Schulman Nathan Linial, Béla Bollobás Miklós Ajtai, Endre Szemerédi Joel Spencer, Fan R. K. Chung Ronald L. Graham, David Avis Noga Alon, László Lovász Shlomo Moran, Richard Pollack Michael E. Saks, Shmuel Zaks Peter Winkler, László Babai |
| filtered | 14 | 1236 | Boris Aronov, Daniel J. Kleitman János Pach, Leonard J. Schulman Nathan Linial, Béla Bollobás Miklós Ajtai, Endre Szemerédi Joel Spencer, Fan R. K. Chung Ronald L. Graham, David Avis Noga Alon, László Lovász Shlomo Moran, Richard Pollack Michael E. Saks, Shmuel Zaks Peter Winkler, Prasad Tetali László Babai |
| factorial k -core | 13.9 | 2678 | Boris Aronov, János Pach Leonard J. Schulman, Nathan Linial Miklós Ajtai, János Komlós Endre Szemerédi, Fan R. K. Chung Ronald L. Graham, Noga Alon László Lovász, Zoltán Füredi Vojtech Rödl, Shlomo Moran Andreas Blass, Richard Pollack Michael E. Saks, Shmuel Zaks |

Table V

THE K-CORES FOR P. ERDOS FOR ALL THE GRAPH VERSIONS

that almost uniformly (except the case of Scheck – 7.43) contributing to his vertex fractional degree. Finally Tanenbaum with a vertex fractional degree 13.0 has a rather small collaboration community with main collaborators Maarten van Steen (contributing a weight 4.68) and Robbert van Renesse (5.4) while the rest is uniformly distributed to the others.

V. CONCLUSIONS

Large and evolving graphs constitute an important element in current large scale information systems. Common cases of such graphs are the Web graph, social networks, citation graphs, CDRs (call data records) where nodes (featured with attributes - in some cases with a large cardinality) are connected to each other with directed edges representing a relation such as endorsement/recommendation/friendship. Community detection and evaluation is an important task in graph mining. A variety of measures have been proposed to evaluate the quality of such communities. In this paper, we evaluated communities based on the k -core concept, as means of evaluating their collaborative nature - a property not captured by the single node metrics or by the established community evaluation metrics. Based on the k -core, which essentially measures the robustness of a community under degeneracy, we extended it to weighted graphs, devising a novel concept of k -cores on weighted graphs. We applied the k -core approach on large real world graphs – such as DBLP and report interesting results. Our contributions are:

| Author | Core | Size | hop-1 list | |
|-----------------------|-------|------|-----------------------|-------|
| C.H. Papadimitriou | 20.80 | 417 | Michalis Yannakakis | 19.62 |
| | | | Erik D. Demaine | 0.14 |
| | | | Georg Gottlob | 1.00 |
| G.Weikum | 16.30 | 1506 | Hans-Jörg Schek | 7.43 |
| | | | Surajit Chaudhuri | 5.05 |
| | | | Raghu Ramakrishnan | 0.41 |
| | | | Gustavo Alonso | 0.43 |
| | | | Divyakant Agrawal | 0.29 |
| | | | Yuri Breitbart | 1.49 |
| | | | Amr El Abbadi | 0.29 |
| | | | Catriel Beeri | 0.33 |
| | | | Rakesh Agrawal | 0.48 |
| | | | Abraham Silberschatz | 0.17 |
| | | | Gautam Das | 0.70 |
| | | | S. Sudarshan | 0.20 |
| | | | Michael Backes | 0.33 |
| | | | Jennifer Widom | 0.19 |
| | | | David J. DeWitt | 0.19 |
| | | | Stefano Ceri | 0.275 |
| | | | Serge Abiteboul | 0.33 |
| | | | Umeshwar Dayal | 0.17 |
| | | | Michael J. Carey | 0.14 |
| | | | Hector Garcia-Molina | 0.14 |
| | | | Yannis E. Ioannidis | 0.23 |
| | | | David Maier | 0.16 |
| | | | Jeffrey F. Naughton | 0.57 |
| | | | Timos K. Sellis | 0.07 |
| | | | Richard T. Snodgrass | 0.07 |
| | | | Jeffrey D. Ullman | 0.07 |
| | | | Henry F. Korth | 0.23 |
| | | | Beng Chin Ooi | 0.08 |
| | | | Edward A. Fox | 0.09 |
| | | | Divesh Srivastava | 0.53 |
| | | | Krithi Ramamritham | 0.15 |
| | | | Christos Faloutsos | 0.13 |
| | | | Victor Vianu | 0.13 |
| DanSuciu | 0.50 | | | |
| Tanenbaum | 13.0 | 4016 | Maarten van Steen | 4.68 |
| | | | Frances M. T. Brazier | 0.98 |
| | | | Howard Jay Siegel | 0.13 |
| | | | M. Frans Kaashoek | 7 |
| | | | Anne-Marie Kermarre | 0.25 |
| | | | Robbert van Renesse | 5.4 |
| | | | Michael S. Lew | 0.02 |

Table VI

FRACTIONAL CORES AND HOP-1 LIST FOR SELECTED AUTHORS.

- i. a novel metric for evaluating the cohesiveness of communities based on the k -core structure
- ii. an innovative extension of the k -core concept assigning weights on the edges and
- iii. an extended experimental evaluation in the case of the DBLP co-authorship graph that results in very interesting findings.

The findings of the experiments on the DBLP co-authorship graph can be retrieved with the prototype systems available at <http://www.lix.polytechnique.fr/~giatsidis/cores/>. There the user may enter an author's name and retrieve

- i. her k -core or f -core ranking (i.e. the best k -core or f -core she belongs too) and
- ii. the respective hop-1 coauthor community.

There are rich visualization capabilities that are enabled covering single authors or chains of co-authors. We capitalized on the graph presentation libraries [2] and [1].

ACKNOWLEDGMENT

Prof. Vazirgiannis is partially supported by the DIGITEO Chair grant LEVETONE in France

REFERENCES

- [1] <http://thejit.org/>.
- [2] <http://www.graphdracula.net/>.
- [3] M. Charikar. Greedy approximation algorithms for finding dense components in a graph. In *Proceedings of the Third International Workshop on Approximation Algorithms for Combinatorial Optimization*, APPROX '00, pages 84–95, London, UK, 2000. Springer-Verlag.
- [4] P. Erdős. On the structure of linear graphs. *Israel J. Math.*, 1:156–160, 1963.
- [5] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.
- [6] S. Fortunato. Community detection in graphs. *Phys. Rep.*, 486(3-5):75–174, 2010.
- [7] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Extracting large-scale knowledge bases from the web. In *Proceedings of the 25th International Conference on Very Large Data Bases*, VLDB '99, pages 639–650, San Francisco, CA, USA, 1999. Morgan Kaufmann Publishers Inc.
- [8] D. W. Matula. A min–max theorem for graphs with application to graph coloring. *SIAM Reviews*, 10:481–482, 1968.
- [9] M. McGlohon, L. Akoglu, and C. Faloutsos. Weighted graphs and disconnected components: patterns and a generator. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08, pages 524–532, New York, NY, USA, 2008. ACM.
- [10] J. Moody and D. R. White. Structural cohesion and embeddedness: A hierarchical concept of social groups. *American Sociological Review*, 68(1):pp. 103–127, 2003.
- [11] S. Papadimitriou, J. Sun, C. Faloutsos, and P. S. Yu. Hierarchical, parameter-free community discovery. In *Proceedings of the European conference on Machine Learning and Knowledge Discovery in Databases - Part II*, ECML PKDD '08, pages 170–187, Berlin, Heidelberg, 2008. Springer-Verlag.
- [12] B. Pittel, J. Spencer, and N. Wormald. Sudden emergence of a giant k -core in a random graph. *J. Combin. Theory Ser. B*, 67(1):111–151, 1996.
- [13] M. Sozio and A. Gionis. The community-search problem and how to plan a successful cocktail party. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, pages 939–948, New York, NY, USA, 2010. ACM.
- [14] G. Szekeres and H. S. Wilf. An inequality for the chromatic number of a graph. *J. Combinatorial Theory*, 4:1–3, 1968.
- [15] S. Wasserman and K. Faust. *Social Networks Analysis: Methods and Applications*. Cambridge: Cambridge University Press., 1994.