Identifying the Presence of Communities in Complex Networks Through Topological Decomposition and Component Densities

Faraz Zaidi, Guy Melançon

Identifying the Presence of Communities in Complex Networks Through Topological Decomposition and Component Densities

Faraz Zaidi*, Guy Melançon*

*CNRS UMR 5800 LaBRI & INRIA Bordeaux - Sud Ouest
351, cours de la Libération
33405 Talence cedex, FRANCE
{faraz.zaidi , guy.melancon}@labri.fr

Abstract. The exponential growth of data in various fields such as Social Networks and Internet has stimulated lots of activity in the field of network analysis and data mining. Identifying Communities remains a fundamental technique to explore and organize these networks. Few metrics are widely used to discover the presence of communities in a network. We argue that these metrics do not truly reflect the presence of communities by presenting counter examples. This is because these metrics concentrate on local cohesiveness among nodes where the goal is to judge whether two nodes belong to the same community or vise versa. Thus loosing the overall perspective of the presence of communities in the entire network. In this paper, we propose a new metric to identify the presence of communities in real world networks. This metric is based on the topological decomposition of networks taking into account two important ingredients of real world networks, the degree distribution and the density of nodes. We show the effectiveness of the proposed metric by testing it on various real world data sets.

1 Introduction

Most real world systems take the form of networks where a set of nodes and edges might be used to represent these networks. Examples include social networks, metabolic networks, world wide web, food web, transport networks. Community detection remains an important technique to organize and understand these complex networks (Girvan and Newman (2002)). Roughly speaking, we like to define a community as a decomposition of a set of entities into 'Natural Groups'. Detection of communities has a wide range of applications in various fields. For example, in social networks, community detection could lead us towards a better understanding of how people collaborate with each other or in a transport network, a community might represent cities or countries well connected through transportation means.

Broadly speaking, research in the field of network analysis can be divided into two categories. First, by developing some metrics that can help us to analyze and detect community structures and second, developing algorithmic procedures to find and group the communities present in the networks. In this paper, we focus on different metrics proposed for community

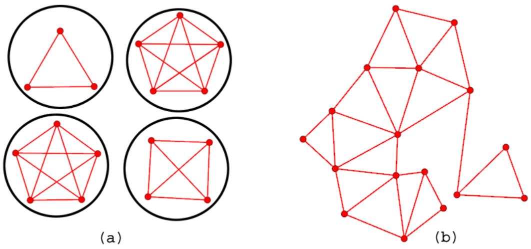

FIG. 1 – *Two networks with the same number of nodes and edges (a) Encircled nodes represent four perfect community structures disconnected to each other and densely connected within and (b) represents one single connected component with nodes sharing neighbors.*

detection. The motivation of this work comes from a simple yet important question: Is there a metric that can assure me the presence of communities in a network?

To answer this question, we need a formal definition of a community. Until now, we avoided using the term *clustering* which is probably a more generic formalism of the term *community*. This is because of a metric called *clustering coefficient* which might be misleading, we will discuss this metric further in Sec. 2. Sociologists use the term *community* (Coleman (1964)) as compared to the statistical and data mining domain where people use the term *clusters* (Tryon (1939)) to refer to the same concept. Thus a community or a cluster can be defined as a group of elements having the following properties as described by Wasserman and Faust (1994):

- Mutuality: Group members choose each other to be included in the group. In a graph-theoretical sense, this means that they are adjacent.
- Compactness: Group members are well reachable for each other, though not necessarily adjacent. Graph-theoretically, elements of the same cluster have short distances.
- Density: Group members have many contacts to each other. In terms of graph theory, that is group members have a large neighborhood inside the group.
- Separation: Group members have more contacts inside the group than outside.

Based on these concepts, a perfect community structure in a network would be represented by disconnected cliques as shown in Fig. 1(a). The connected components in Fig. 1(a) satisfy all the properties described above as being a perfect community structure. Notice that the two graphs has exactly the same number of nodes and edges. To the best of our knowledge, there is no such metric which tries to identify the presence of communities in a network by analyzing the graph on the whole in a global perspective. There are metrics like Clustering Coefficient and Jaccard Index (see section 2 for more details) that determine local cohesiveness of a set of nodes, i.e. they focus on the immediate neighborhood of nodes but they fail to capture the presence of communities on the whole as argued by different researchers (Brandes and Erlebach (2005); Girvan and Newman (2002)). Fig. 1(b) is an example that depicts this phenomena where several nodes share common neighbors but it is difficult to identify communities in such a network.

In this paper we aim to develop a metric which helps us to identify these dense components as shown in Fig. 1(a). We use two important ingredients of real world networks; the *node* *degree* to decompose the network into several subgraphs and *density of nodes* to identify the presence of a community. Being able to identify communities in a network has several real world applications such as grouping similar entities into groups can help us understand and model the behavior of real world systems. Once we identify the presence of communities in a network, eventually an algorithm can be built to group these communities. But in this paper, we limit ourselves to proposing only a metric where a detailed study of a possible clustering procedure and comparing its results with other clustering algorithms remains out of scope.

A huge advantage of the proposed metric is that it can be calculated in almost linear time making it applicable on large size data sets as compared to different indices that have very high run time complexity. Moreover the topological decomposition implicitly embeds the knowledge of node degrees which helps us to identify the presence of communities with respect to the degree of nodes.

We briefly describe different metrics present in the literature related to the community detection problem in section 2. Section 3 presents the different data sets used for experimentation. In section 4 we present the mathematical details of the proposed metric. Section 5 is devoted to analyzing different data sets with respect to the proposed metric followed by interesting observations derived from the proposed metric in section 6. Finally we present conclusions and future research directions.

2 Related Work

Different metrics exist in the literature to study the community detection problem. Broadly classifying these metrics, we can say that some metrics are local i.e. calculated upon either nodes or edges as compared to metrics that are calculated over the entire graph. An example of a local metric would be the in-out degree of a node. Our goal is to find a metric that is calculated for the entire graph and does not focus on individual nodes and edges. On the other hand, examples of popular global metrics include Betweenness Centrality and Node Eccentricity (Brandes and Erlebach (2005)). To the best of our knowledge, none of the global metrics are designed to identify the presence of community structures. Below, we review a few metrics that address the community detection problem locally.

One of the most widely used metric in network analysis is the clustering coefficient proposed by Watts and Strogatz (1998). This metric can be often misleading due to its name as this metric does not guarantee the presence of communities in a network. For example, consider the two networks shown in Fig. 1. Calculating the average clustering coefficient of network (a) gives us a value of 0.88 and that of network (b), a value of 0.61. Both these values do not reflect the underlying structure of the network where Fig. 1(a) has intuitively distinct community structure as compared to Fig. 1(b). This metric along with the average path length were used by Watts and Strogatz (1998) to classify networks as being small world networks having 'six degree of separation' principle (Milgram (1967)).

Another popular metric is the *Jaccard Index* introduced by Jaccard (1901) also known as *Jaccard similarity coefficient*. This metric is used to measure the similarity of two elements based on common neighborhood. More precisely the index looks at the number of common neighbors of the two elements and compares it with the size of all the neighbors of the two elements. An edge is assigned high similarity value if they share lots of neighbors. Coming back to our example in Fig.1(a), if we consider the edges connecting the clique with three nodes only, all its edges are assigned a value 0.33 as compared to the edges of the clique with five nodes that are assigned a value 0.6. A low edge value might suggest that the an edge is not part of a community which in this case, is contradictory to the definition of a community structure we presented earlier, as the edge with 0.33 value is in fact part of a community.

Several Other metrics have been proposed in the literature where Melançon and Sallaberry (2008) provide a good comparative study of various local metrics for the community detection problem. The jaccard index clearly stands out as the archetype metric for finding communities in networks based on the notion of a triad. Readers are recommended (Melançon and Sallaberry (2008)) for further details.

We would like to refer to another metric which is not used for community detection, rather to classify networks and has gained a lot of popularity in the domain of network analysis. The metric, degree distribution classifies networks as being scale free if the degree distribution follows a power-law. In other words, this means that there are a few nodes that have a very high number of connections and lots of nodes have only a few connections. These networks have many interesting properties (Barabasi and Albert (1999)) and we use this knowledge to get inspiration for our metric where details are described in section 4.

3 Data Sets

We use several real world data sets for experimentation.

Co-authorship Network contains the collaboration network of scientists posting preprints on the condensed matter archive at `www.arxiv.org` between 1995-1999, as compiled by Newman (2001). The network contains 1670 nodes and 47600 edges.

Movie Database is an actor-actor graph where two actors are connected to each other if they have acted in a movie together. It contains a reachable graph of distance 5 starting from the actress Sharon Stone from movies made until the year 1999 (see Archambault et al. (2007) for more details). This network contains 7640 nodes and 27600 edges approx.

Air Transport Network is an airport-airport graph where edges represent a direct flight from one airport to the other. This network has attracted lots of researchers from the field of geography and transportation (see Rozenblat et al. (2008) for details). This network contains 1540 nodes and 16500 edges.

Internet Tomography Network is a collection of routing paths from a test host to other networks on the Internet. The database contains routing and reachability information, and is available to the public from the Opte Project website (`http://opte.org/`). This network contains 35800 nodes and 42400 edges.

American Football contains the network of American football games between Division IA colleges during regular season Fall 2000, as compiled by Girvan and Newman (2002). Teams are represented by nodes and edges represent a game between two teams. This network contains 115 nodes and 616 edges.

To further assert the efficiency of our metric, for each of these real world networks, we generate artificial networks of equal sizes using two known graph models, the random networks (Erdos and Renyi (1960)) and the small world network (Watts and Strogatz (1998)). We expect to find community structures in small world networks where as an absence of communities from the random networks.

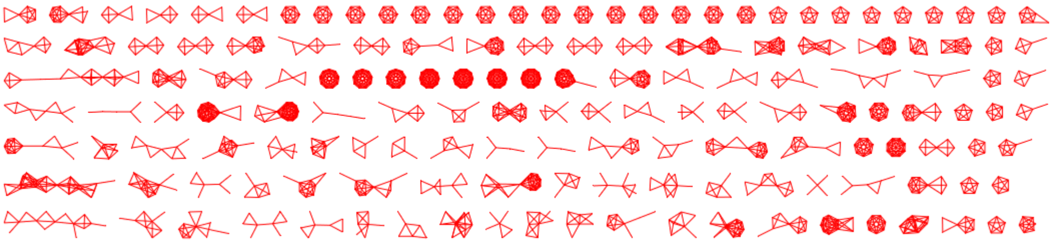

FIG. 2 – *A partial view of Max$_{12}$−DIS of Co-authorship network where disconnected components can be easily identified for high density.*

4 Topological Decomposition for Edge Distribution

As discussed previously, in this paper, we present a new metric to identify the presence of communities. The goal is obviously to find densely connected nodes that can be identified as communities. Our inspiration comes from the fact that in real world networks, the degree distribution is not random, rather it follows a pattern where different nodes have varying degrees (Barabasi and Albert (1999)). In the presence of nodes having high degree, it is difficult to identify dense components in a network (Zaidi et al. (2009)). Thus as a first step, we introduce a decomposition based on the topology of the network. As a result, the network breaks into several components disconnected from each other as shown in Fig.2. We can then calculate the densities of each connected component to identify the presence of dense components in the entire network.

Thus we propose a three step approach, topological decomposition, connected component identification and calculating component densities. All these steps can be performed in linear time in terms of number of edges in the worst case considering that the maximum degree in a graph is bounded by a low constant factor.

Topological Decomposition: We introduce the idea of Max$_d$-Degree Induced Subgraph (Max$_d$-DIS) where Max$_d$-DIS is an induced subgraph constructed by considering only the nodes having degree at most d in G. Mathematically for a graph $G(V, E)$ where V is a set of nodes and E is a set of edges, the Max$_d$-DIS is defined as $G'(V', E')$ such that $V' \in V$ and $E' \in E$ and $\forall u \in V', Deg_G(u) \leq d$ where d can have values from 0 to the maximum node degree possible for a network. Similarly we can construct Min$_d$-DIS where $Deg_G(u) \geq d$. Through out the paper, we use the term DIS to refer to both max and min degree induced subgraphs.

Lets consider an example from the Co-authorship network by drawing a Max$_{12}$−DIS. Note that a node having degree 12 in the original graph might not have the same degree in the induced subgraph. Moreover it is also evident that these nodes might no longer be connected to each other as shown in Fig.2. By construction, when we consider a Max-DIS, we essentially include nodes of up to degree d only. This helps us to understand how the edges are distributed among nodes having degree $d - 1$.

From the Fig.2, it is evident the Max$_{12}$−DIS helps us to identify dense components of up to degree $d - 1$ in the entire network. Consider part of a graph where a clique of 6 nodes exists

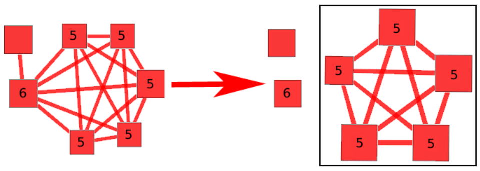

FIG. 3 – *Example of Max$_5$−DIS before and after calculating Max$_5$−DIS*

as shown in Fig.3. The clique is connected to the entire network through a single node. All the nodes in this clique have degree 5 except the node which connects this clique to the entire network, which in this case has degree 6. When we build the Max$_5$−DIS the node with degree 6 will not be included in the induced subgraph and only the enclosed nodes having degree 5 will be part of the Max$_5$−DIS. Thus in this way, we can identify densely connected nodes in the entire network.

We argue that iterating over the Max$_d$-DIS from small values of d to the maximum possible degree in the original network, we can identify the presence of densely connected components, if there are any. Creating induced subgraph for all possible values of node degrees is bounded by the maximum degree possible in the entire network. Since we are going to calculate the metric for both Max$_d$-DIS and Min$_d$-DIS, the time complexity of creating the induced subgraph is $O(2*max_d*m)$ where max_d is the maximum node degree and m is the total number of edges in G .

Calculation of Connected Components: The next step is to calculate all the connected components in the subgraph. We use a breadth first search algorithm (BFS) starting from a node and iterating through its neighbors to find the connected component it belongs to. Once we have identified nodes connected to the start node, we restart the BFS from a node that has not yet been visited. All the connected components of a graph can thus be calculated in $O(n+m)$ time where n is the number of nodes and m is the number of edges in G.

Measuring Component Densities of Graphs: Now that we have decomposed the graph and identified connected components, we calculate a metric to quantify the presence of densely connected components if there are any. We use a metric proposed by Watts and Strogatz (1998) called local clustering coefficients but since we apply it to individual connected components as opposed to the whole graph, we prefer to call it component densities. We assign a component density to each individual connected component using the following equation.

$$CD_k = (e_k * 2)/(n_k(n_k - 1)) \tag{1}$$

Where CD_k represents the Component Density (CD) of connected component k, e_k represents the number of edges in k and n_k represents the number of nodes in k. The equation represents the ratio of the number of edges in component k to the maximum number of edges possible in the component. A value of 1 suggests that the component is a clique and since the component is connected the minimum CD value possible is $2/n_k$. Note that checking whether a connected component is a clique is no more than a counting problem where we can identify the presence of a clique by simply counting the number of nodes and the number of edges in a connected component. We would like to mention that we do not address the well known maximal clique problem using this metric which is known to be NP-Complete (Cook (1971)). Moreover we do not try to find cliques of a fixed size *k* which is shown to be solvable in $O(n^k)$ by Nesetril and Poljak (1985) as our the method does not gurantee that we will find cliques of some fixed size *k*. The proposed method is capable of finding cliques in linear times in terms of number of edges irrespective of the size of the clique.

Once we have calculated the CD for individual connected components in the DIS, we calculate the weighted component density for all the connected components in a DIS. We represent this value by CDG$_d$ for d degree induced subgraph and is calculated by the equation:

$$CDG_d = \sum_{k=0}^{kmax} \frac{n_k * CD_k}{n_d} \tag{2}$$

Where CDG$_d$ represents the weighted Component Density of d-DIS. While calculating the CDG$_d$, we exclude the weight of components having only 1 or 2 nodes as it biases the CDG$_d$. We do count them in the total number of nodes (n_d) present in the induced subgraph though. This is because if a graph has lots of 0 degree nodes, its overall component density will be lower than a graph with well connected higher degree nodes. The weight is associated to ensure that components having more nodes are weighted more as compared to components having fewer nodes. The CDG$_d$ can be calculated for different values of d where the d can take values from 0 to the maximum degree of a node in G. The calculations in Eq. 1 and 2 are totally independent of how the subgraph was constructed and thus can be used to calculate the component densities of either Max$_d$-DIS or Min$_d$-DIS.

The CDG values of Max$_d$-DIS (given by CDG$_{Max_d}$) and Min$_d$-DIS (given by CDG$_{Min_d}$) represents the presence or absence of dense components in G. High values of CDG suggest that there are densely connected components in the induced subgraph which eventually represent communities in G. Another supplementary information that can be extracted from CDG values is that by identifying the peak value of CDG$_{Max_d}$ and CDG$_{Min_d}$, we can point the out the induced subgraphs in which these communities are present as shown in Fig.2.

We plot graphs for the respective CDG$_{Max_d}$ and CDG$_{Min_d}$ values for the data sets described in section 3 as shown in Fig.6. The values on the x-axis represent the maximum degree of each network, which in turn represents the number of subgraphs generated for each network. The values on the y-axis are the CDG values which can be between 0 and 1 where 1 represents the presence of perfect community structure with cliques. For each real world data set, we have also generated small world and random networks of equivalent number of nodes and edges so that we can compare the behavior of the metric with the corresponding artificially generated network. These networks are drawn using Dotted-Line for Small world networks and Dashed-Line for Random networks.

5 Analyzing Different Data Sets using Components Densities of Graph

The evaluation of CDG$_{Max_d}$ and CDG$_{Min_d}$ for Random networks and Small world networks for graphs of different sizes can be generalized easily. For random networks, we do not expect to find any community structure and thus these networks have low CDG values for all the test cases as shown in Fig.6. On the other hand, we have the small world networks which by definition contain communities and this is well reflected by the high CDG values for all the artificially generated networks. One exception is the case where the generated small world graph is equivalent to the size of the Internet network. This is due to the low overall density of the Internet graph itself as the edges scale linearly with the number of nodes.

We first look at the **Co-authorship** and the **Movie** networks. Both these graphs have similar CDG$_{Max_d}$ and CDG$_{Min_d}$ values as shown in Fig.6(a)(b)(c)(d). This is due to the fact that both these networks follow the small world and scale free structure where there are many low degree nodes and a few nodes dominated most number of connections representing the scale-freeness of the graphs. This phenomena can be deduced from (Fig.4(a)) where we show the plot for the degree distribution of the Movie network. Moreover the high CDG$_{Max_d}$ values suggest that there are communities present in the low degree nodes as well as in the high CDG$_{Min_d}$ values representing the small world architecture of the two networks.

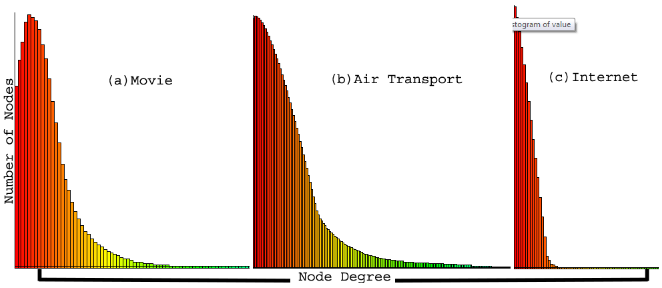

FIG. 4 – *Degree distribution for the Movie, Air Transport and the Internet network.*

In case of the co-authorship network (Fig.6(a) and (b)), the peak values of 0.4 for CDG$_{Max_d}$ are attained when the degree values are between 6 and 8. This suggests that most of the dense communities are present when node degrees have a maximum value of 8. This result has logical semantics to it, as we are considering a collaboration network of researchers and they are connected to each other if they publish an article. Mostly 6 to 8 people appear as authors in an article forming cliques in the collaboration network. This information is well represented by the CDG$_{Max_d}$ values. On the other hand, the CDG$_{Min_d}$ represents the collaboration of authors who publish a lot. Since the data set comes from a particular domain of research, it is not a surprise to see that the people having the highest degrees collaborate with each other. The Movie Database follows a similar structure to that of the Co-authorship network. The peak value for CDG$_{Max_{16}}$ is 0.87. Drawing the Max$_{16}-$DIS, we can clearly see different actors are densely connected to each other as shown in Fig.5(a).

The **Air transport** network is an interesting example(Fig.6(e) and (f)). It has a scale free structure as can be seen from the degree distribution of the network in Fig.4(b). This network is not classified as a small world network as it does not have a high clustering co-efficient (Rozenblat et al. (2008)). Using our metric, we are still able to find densely connected nodes as we get high values for Min$_d-$DIS. The community structure found is shown in Fig.5(b) where all the worlds busiest airports are connected to each other through a direct flight. This makes

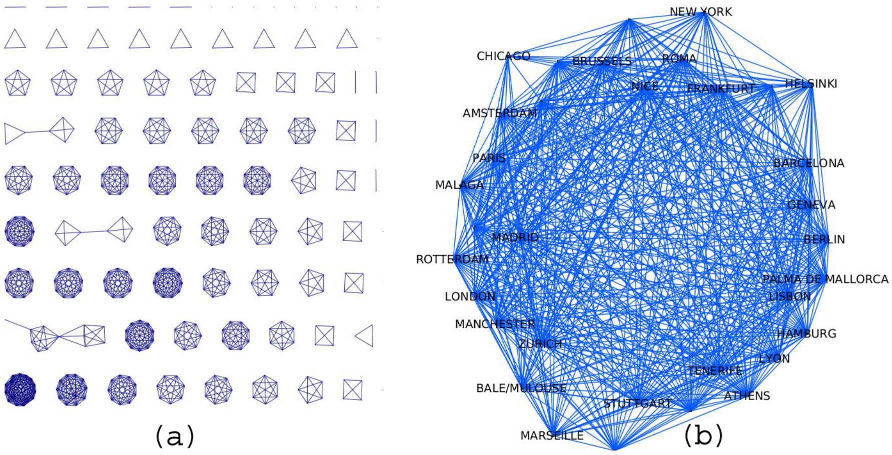

FIG. 5 – *(a) Max$_{16}$−DIS for the Movie database representing lots of densely connected components (b) Min$_{158}$−DIS for the Air Transport network where all the major cities in the world are well connected to each other*

sense as the worlds most important airports like New York, Paris, London all have a direct flights to each other.

The **Internet** network is not classified as a small world as it is has low clustering coefficient. Using our metric, we do not find any community structure neither in the CDG$_{Max_d}$ nor in the CDG$_{Min_d}$ graphs thus reflecting the correctness of the metric. The degree distribution of the internet graph is shown in Fig.4(c).

Finally the **Football** network where we suspected an absence of communities within this network as teams might like to uniformly distribute the number of games amongst their opponents which turned out to be a false assumption. It is quite clear from the high CDG$_{Max_d}$ and CDG$_{Min_d}$ that there are communities present in this network. This is because certain teams play more frequently within the same region or by playing with the teams that have a history behind them. An important remark using the CDG$_{Min_d}$ values is that the highest degree nodes do not form a community structures which suggests that teams playing more games not necessarily play against each other.

6 Inferences and Observations

Apart from identifying the presence of communities in different networks, interesting properties of real world networks can be observed using the CDG$_{Max_d}$ and CDG$_{Min_d}$ graphs. We list these below:

- Networks are usually classified as either Random, Small World, Scale Free or both Small World-Scale free at the same time. Using the proposed metric, we can have further insight in these networks by understanding how the edges are distributed in low or high degree nodes of these networks. Communities can exist in nodes that have a low degree in the graph (as is the case with Co-authorship and Movie networks), or they can exist only in high degree nodes (Air Transport network) or in nodes having an average degree (Football network).

- The absence of communities can have two consequences, either the network is Random or it follows a scale free behavior. In case of scale free behavior, networks usually consist of star like structures where many nodes connect to a single node. Typical example is the routing information of servers in case of the Internet network. From the proposed metric, we are able to find the same behavior in the Air transport network where the degree of nodes is not very high. Logically speaking, an airport in a capital city of a small country will play the role of a hub where the smaller cities will have a direct flight to the capital city which will eventually be connected to one of the major airports of the world.
- Another interesting observation about the Internet network is the absence of community structure in the high degree nodes. This means that if two high degree nodes are to communicate with each other, they pass through a low degree node.

Although this preliminary analysis reveals some interesting facts about the different data sets, a more detailed study by the domain experts might reveal more information. We believe that this analysis and any other information can help researchers find better ways to group entities together. Moreover since the analysis is highly efficient in terms of time complexity, it is quite practical for future research where the size of data is growing exponentially.

7 Conclusions and Future Work

In this paper we have introduced a metric based on the topological decomposition of graphs. We call this metric, component density of graphs. The metric helps us to identify the presence of community structure in real world complex networks. Calculation of this metric takes linear time in terms of number of edges and proves to be very fast when applied to large size data sets. We have tested the metric with different data sets and show the effectiveness of the metric by comparing the results with small world and random network models. Moreover the topological decomposition proposed opens new dimensions in the field of visual data mining as complex networks can be simplified using the proposed method. As part of future work, we believe that this metric can lay a foundation for building a high speed clustering algorithm for large size networks. Since this metric is quite efficient, discovering clusters using this metric will remain highly efficient in terms of time complexity.

References

Archambault, D., T. Munzner, and D. Auber (2007). Grouse: Feature-based, steerable graph hierarchy exploration. In *EuroVis*, pp. 67–74.

Barabasi, A. L. and R. Albert (1999). Emergence of scaling in random networks. *Science 286*(5439), 509–512.

Brandes, U. and T. Erlebach (2005). *Network Analysis : Methodological Foundations (Lecture Notes in Computer Science).* Springer.

Coleman, J. S. (1964). *An Introduction to Mathematical Sociology*. Collier-Macmillan, London, UK.

Cook, S. A. (1971). The complexity of theorem-proving procedures. In *Proc. of the 3rd Annual ACM Symp. on Theory of Computing*, pp. 151–158.

Erdos, P. and A. Renyi (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci 5*, 17–61.

Girvan, M. and M. E. J. Newman (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA 99*, 8271–8276.

Jaccard, P. (1901). Bulletin del la société vaudoisedes. *Sciences Naturelles 37*, 241–272.

Melançon, G. and A. Sallaberry (2008). Edge metrics for visual graph analytics: A comparative study. In *IV*, pp. 610–615. IEEE Computer Society.

Milgram, S. (1967). The small world problem. *Psychology Today 1*, 61–67.

Nesetril, J. and S. Poljak (1985). On the complexity of the subgraph problem. *Comment. Math. Univ. Carolinae 26*, 415–419.

Newman, M. E. (2001). Scientific collaboration networks. i. network construction and fundamental results. *Phys Rev E Stat Nonlin Soft Matter Phys 64*(1 Pt 2).

Rozenblat, C., G. Melançon, and P.-Y. Koenig (2008). Continental integration in multilevel approach of world air transportation (2000-2004). *Networks and Spatial Economics*.

Tryon, R. C. (1939). *Cluster analysis.* Edwards Brothers, Ann Arbor, Michigan.

Wasserman, S. and K. Faust (1994). *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press.

Watts, D. J. and S. H. Strogatz (1998). Collective dynamics of 'small-world' networks. *Nature 393*, 440–442.

Zaidi, F., A. Sallaberry, and G. Melançon (2009). Revealing hidden community structures and identifying bridges in complex networks. In *WI-IAT '09: Proceedings of the 2009 IEEE/WIC/ACM Conference*, pp. 198–205.

Résumé

La taille des jeux de données ne cesse exploser dans de divers domaines tels que les réseaux sociaux et internet. Cette explosion a dynamisé les travaux de recherche sur les analyses de réseaux et la fouille de données. L'identification de communautés reste une méthode classique pour explorer et classifier de tels réseaux. Pour celà, quelques métriques, souvent les mêmes, sont fréquement utilisées. Nous allons montrer, grâce à des contre-exemples, que ces métriques ne peuvent pas mettre en évidance l'existance de communautés. Celà parce qu'elles comparent localement la similarité des noeuds, alors que l'objectif est de déterminer s'il existe des communautés en considérant le réseau dans sa globalité. Dans cet article, nous proposons une nouvelle métrique pour identifier la présence de communautés dans des réseaux tirées du monde réel. Cette métrique est basée sur une décomposition topologique des réseaux, prenant en compte deux acpects importants de ces réseaux, la distribution des degrés ainsi que la densité des noeuds. Nous démontrerons l'efficacité de la métrique proposée en l'applicant à de nombreux jeux de données de la vie courante.

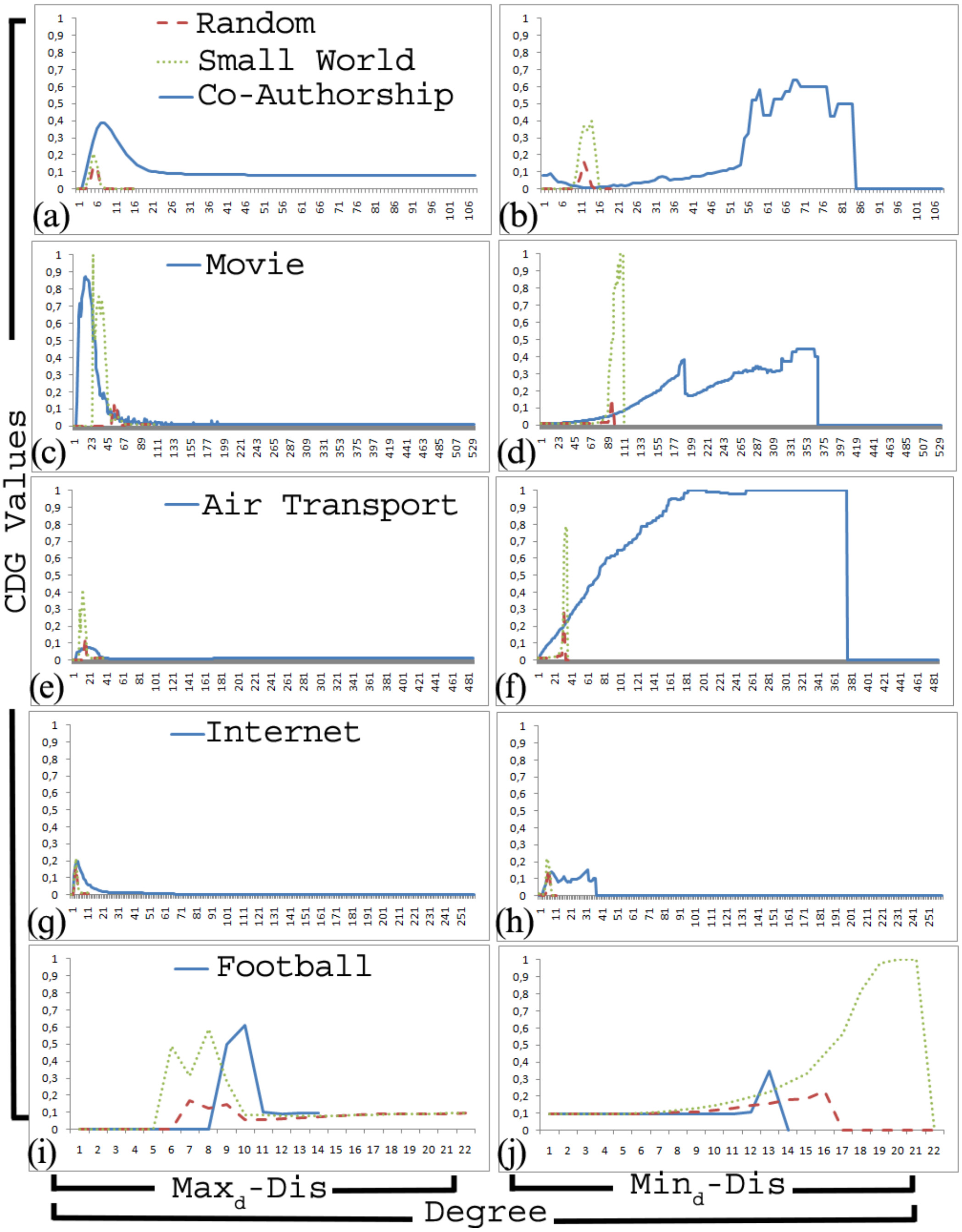

FIG. 6 – *Component Densities of Graphs for the 5 data sets.{(a)(c)(e)(g)(i)} represent the CDG$_{Max_d}$ and {(b)(d)(f)(h)(j)} represents the CDG$_{Min_d}$ values.*