Overlapping Community Detection Versus Ground-Truth in AMAZON Co-Purchasing Network

Malek Jebabli, Hocine Cherifi, Chantal Cherifi, Atef Hamouda

Overlapping community detection versus ground-truth in AMAZON co-purchasing network

Malek Jebabli
University of Burgundy
Dijon, France
Email: Malek.Jebabli
@u-bourgogne.fr

Hocine Cherifi
University of Burgundy
Dijon, France
Email: hocine.cherifi
@u-bourgogne.fr

Chantal Cherifi
University of Lyon 2, France
Email: Chantal.BonnerCherifi
@univ-lyon2.fr

Atef Hamouda
University of Tunis El-Manar
Tunis, Tunisia
Email: atef_hammouda
@yahoo.fr

***Abstract*—Objective evaluation of community detection algorithms is a strategic issue. Indeed, we need to verify that the communities identified are actually the good ones. Moreover, it is necessary to compare results between two distinct algorithms to determine which is most effective. Classically, validations rely on clustering comparison measures or on quality metrics. Although, various traditional performance measures are used extensively. It appears very clearly that they cannot distinguish community structures with different topological properties. It is therefore necessary to propose an alternative methodology more sensitive to the community structure variations in order to conduct more effective comparisons. In this paper, we present a framework to tackle this challenge through a comprehensive analysis of the community structure of overlapping community structured networks. We illustrate our approach with an experimental analysis of a real-world network with a ground-truth community structure that we compare with the output of eight different overlapping community detection procedures, representative of categories of popular algorithms available in the literature. The results allow a better understanding of their behavior. Furthermore, they demonstrate that more emphasis should be put on the topology of the uncovered community structure in order to evaluate the effectiveness of community detection algorithms.**

***Keywords*—Community structure, detection algorithms, overlapping community networks, network analysis.**

I. Introduction

The community detection problem has led to an impressive body of literature, and many community detection methods and surveys have been introduced in recent years. Although, there has been a tremendous effort on introducing new algorithms in order to uncover this hidden structure of a network, little attention has been devoted to various complementary aspects of this issue. First of all, there is no formal consensus on a definition that captures the gist of a community. It is intuitively understood as a cohesive group where members interact with each other more intensely than with those outside the group. As there are many diverse understandings of how cohesiveness translates in formal graph-theoretic terms, community detection has been approached from many different perspectives.

Second, the lack of labelled ground-truth data has limited the understanding of the community structure in real-world networks. Recently, the situation has greatly evolved through the work of Yang and Leskovec [1]. The authors identified a set of large-scale real-world networks where a functional notion of ground-truth communities can be defined. In other words, nodes can be explicitly classified in diverse meaningful groups. With these data, it is therefore possible to gain a better knowledge about the topological properties of community structure. However, there is no guarantee that communities defined on a functional basis are encoded in the structural information of the network.

And last but not the least, very few attention has been devoted to the evaluation issue. Indeed, it is essential to compare the effectiveness of the various community detection algorithms. This complex and open problem is classically considered either from the clustering perspective or from the quality perspective. When there is a ground-truth of the community structure, validation is simply accomplished by comparing discovered communities against known ones. Various clustering-comparison measures have been proposed that can be classified into three main categories: measures based on pair counting, set-matching-based measures and information-theoretic-based measures. With pair-counting-based measures, clustering comparison is based on counting the pairs of points on which two partitions agree or disagree. The Rand Index (RI)[2] and the Jaccard Index are well-known measure in this class. It should be remembered that, there are many other measures in this class [3]. However, after correction for chance, many of these measures are equivalent [4]. Set-matching-based measures are based on set cardinality. They intend to find the largest overlaps between pairs of different partition clusters. Purity is the proportion of correctly assigned nodes. Each identified cluster is matched to the one with the maximum overlap in the reference cluster, and then the accuracy of this assignment is measured by counting the number of correctly assigned nodes. Information-theoretic-based measures have gained increasing attention in the clustering literature. They are based on the mutual information shared by two partitions in order to assess their agreement. Normalized mutual information (NMI) is defined as the ratio of the mutual information to the mean value of the entropy of both partitions. It takes the value of 1 when the two partitions are identical and 0 when they are independent. When the underlying community structure is unknown, quality functions are used. They are based on various properties that can be encountered in a "good" community structure. Most of them formalize in different ways the idea that communities are sets of nodes densely connected and poorly connected to the rest of the network. Modularity is the most widely used quality function to compare the effectiveness of community-detection algorithms on real data with unknown community structure. It expresses how a community structure

has a high-density ratio as compared to a random graph with the same degree sequence. The main weakness of the quality function approach is that very often the detection algorithms also use it as an optimization criterion, therefore it introduces a bias in the comparisons. Furthermore, one needs to be very cautious in order to define what is a good structural property.

Some recent works points out the weakness of these approaches [5], [6]. Indeed, there are a wide variety of community detection algorithms whose output exhibits high structural variability. Relying only on global clustering or quality measures do not allow to reveal nuances of the structure of real and extracted communities because these metrics ignore the topology of the community structure. Indeed, two estimated community structure can reach a close level of NMI performance while their link distributions are quite different.

As pointed out by these authors, it is necessary to analyze the topological properties of the community structure. Our work is in this line. In order to compare the efficiency of overlapping community detection methods, we propose to analyze the topological properties of their outputs. To do so, starting from the uncovered community, we build and analyze the topological properties of the community network. In this network, the nodes are the extracted communities and there is a link between two nodes if the two communities overlap. The basic idea of the proposed approach is that an efficient community detection algorithm must be able to output a community structure with comparable topological properties than the one obtained with the ground truth community structure. Even if some nodes are misclassified, or if its structure does not agree with what is generally expected trough some quality measures, it is important that it can encode the topology of the community network.

To illustrate our framework with an experimental analysis, we use the AMAZON network of product co-purchases and the most frequently used overlapping community algorithms. We construct the ground truth network of community as well as the estimated networks of community (using the community structure given by the overlapping community algorithms). To investigate the community structure properties, we consider topological measures from different scales. At the macroscopic level, we compute the average clustering coefficient, average shortest path, diameter, density and degree correlation. At the microscopic level, we accurately analyze the distribution of the node degree, the average clustering coefficient as a function of degree as well as hop distance. Finally, we study a mesoscopic property of the community structure, namely the community size distribution.

Note that our main focus is not to provide another empirical comparative evaluation of overlapping community detection algorithms, but rather at pressing the point on the importance of the topological aspect of these comparisons in order to ascertain a clear picture of the effectiveness of the algorithms.

The rest of this paper is divided into four sections. Section 2 discusses some related works. Section 3 presents the dataset and the community detection methods. In section 4, we report analysis results with some discussions. Finally, section 5 summarizes our concluding remarks.

II. Related work

The paper by Xie et al. [7] is one of the most influential contributions to the overlapping community detection problem. The authors present a thorough comparison of fourteen algorithms on numerous artificial and real-world networks using various performance measures. They use synthetic networks generated using the LFR (Lancichinetti, Fortunato & Radicchi) model proposed in [8] in order to study the behavior of the various community detection algorithms. This model provides a rich set of parameters to control the network topology. Furthermore, it allows generating synthetic networks with features close to the ones observed in real-world networks. Its main drawback is that this model requires that all the overlapping nodes interact with the same number of communities, which is quite unrealistic in practice. Extensive comparisons have been conducted over different overlapping densities and community size ranges. The quality of the detected communities is measured through the NMI and the Omega-Index [9], which is the overlapping version of the Adjusted Rand Index (ARI). Four limiting cases are considered (small or large community size ranges together with low or high overlapping density). To summarize the results, average ranking scores are computed for both measures. Results suggest that NMI and Omega provide similar overall evaluation to some extent. Averaging the scores across the four different cases, it appears that the top seven algorithms are exclusively agent-based algorithms. In order to provide insight into the behaviors of different algorithms, the community sizes distribution is compared with the known ground truth. The authors conclude that observations on the community size distribution can be used to explain the ranking of the algorithms. In other words, when the discovered distributions are in agreement with the ground truth distribution, the algorithms perform well with respect to ranking and conversely. They also analyze the algorithm's ability to detect the overlapping nodes using the F-score as a measure of detection accuracy. Results of the experiments allow to clearly identify the algorithms that tend to over or under-estimate the overlapping nodes. Furthermore, the rankings with respect to F-scores provide quite different pictures of the performance as compared to NMI or Omega score. Indeed, these two types of rankings provide complementary information. NMI and the Omega index measure detection performance at the community level, while F-measure focuses at the node level. One of the main lessons of this work is to highlight the complexity of the performance measures issue. Indeed, measures like NMI and Omega focus only on providing an overall measure of algorithmic accuracy, and complementary measures are needed to perform a more precise analysis. Orman et al. [5] present a comparative study of a representative set of community detection algorithms using the LFR model with appropriate parameters to generate realistic undirected non-overlapping networks. They evaluate the outputs of the community detection algorithms trough classical partition based performance measures together with community structure topological measures. It turns out that the partition based measures (Purity, RI, ARI and NMI) agree with each other with small differences when considering the way they rank the algorithms. Arguing that the main drawback with these measures is that communities are compared only in terms of individual node membership, without taking the underlying topology into account, they propose to use

community-oriented topological measures (embeddedness, internal degree, community size, internal transitivity, scaled density, average distance, and hub dominance). Results of their experiments show that extreme ranking (worst and best) of the algorithms are similar for both types of measure. However, this is not the case for the other algorithms. Indeed, partition-based measures performance can be relatively high, although the uncovered communities substantially differ from the reference, according to the topological measures. They conclude that both approaches are complementary and must be used in order to perform a relevant and complete analysis of community detection results.

In a recent survey, Arenberg et al. [10] evaluate four overlapping community detection algorithms together with nine non-overlapping ones on large-scale real-world networks. The networks, with overlapping ground-truth communities, are available in the Stanford Large Network Dataset Collection (Amazon, DBLP, Livejournal, Orkut, Youtube). For each network, the top 5000 ground-truth communities (with the bigger size) are ranked based on internal density and the bottom quartile is removed. In addition, five more graphs with disjoint ground-truth communities are derived from the real-world networks in order to test the algorithms for disjoint communities. In this study, the algorithms are compared using quality measures as well as clustering-based measures. Four quality metrics has been chosen among the various quality functions available in the literature [11] (internal density, clustering coefficient, conductance, and triangle participation ratio). Several statistical measures (precision, recall, F-measure, specificity, accuracy, NMI) based on the confusion matrix are used to measure the similarity between the set of ground-truth communities and the set of communities output by an algorithm. Results of their investigations show that these two types of measures are not equivalent. An algorithm that identifies communities with good structural properties does not necessarily yield good clustering-based performance metrics. It may be pointed out that according to Yang and Leskovec study, many quality metrics are highly correlated [11]. Indeed, it appears that the eleven quality metrics they investigate in their paper can be grouped into four clusters. Note that internal density and triangle participation ratio belong to the same cluster, and therefore return highly correlated values.

In [12] Fortunato et al. compare the community structure uncovered by ten popular community detection algorithms on a collection of real-world and synthetic networks. Three of the algorithms produce non-overlapping communities where the others allow overlaps. The fifteen real networks come with known ground truth communities and the synthetic networks are generated using the LFR model. Real-world networks can be classified in two groups. The first group is made of small-size networks classically used as testbed in the community detection literature (Zachary, football, etc.), while the second group contains more recent and challenging large-scale networks such as Amazon and Livejournal. Nine of these networks come with overlapping ground truth communities while the remaining ones possess a non-overlapping community structure. Comparisons are carried out using clustering based measures. Overall, results can be separated into three groups by descending order of NMI scores. The first group (with the highest performances) corresponds to the LFR benchmark. The second group consists of the small-sized classical datasets (karate, football, polblogs, polbooks), and the third group is made of the remaining networks. Globally, in these large-scale networks, uncovered communities do not align well with the ground-truth communities. Restricting the comparison to communities of comparable topological properties (size, link density or embeddedness) does not reveal major improvements. Globally, communities estimated by the algorithms do not match well the ground truth communities and the results are more influenced by the network than by the specific method adopted.

There are a few messages coming from this study. First of all, it raises the question of the community definition. Until now, there is an ambiguity between structural communities as revealed by the topology and functional communities where nodes are grouped in different classes corresponding to their intrinsic features. As the authors report "the field has been silently assuming that structural communities reveal the non-topological classes". Second, relying on the classification of the nodes in order to characterize the community detection algorithms is not sufficient. A detailed investigation of the topological properties of the community structures must be carried out in order to assess the structural dissimilarity among the outputs of community detection algorithms.

III. Dataset and community detection methods

A. Dataset

The network we consider is the AMAZON product co-purchasing network available at http://snap.stanford.edu/data. Nodes represent products and links connect commonly co-purchased products. Each product belongs to one or more hierarchically organized product categories that we view as a ground-truth communities. The basic properties of this network are reported in Table I. Note that, it exhibits the typical characteristics of real-world complex networks. Indeed, with an average shortest path equals to 2.78, it satisfies the small-world property i.e. most nodes are just a few edges away on average. Moreover, it is characterized by a high clustering coefficient. Its transitivity value above 0.2 reflects the tendency of link formation between neighbouring nodes. With a degree correlation value of -0.06, it is dissortative. In other words, highly connected vertices tend to connect to those with few connections.

TABLE I. Global properties of AMAZON network. The calculated properties are number of nodes (Nodes), number of edges (Edges), Density, Average Shortest Path (ASP), Assortativity Coefficient (AC), and Clustering Coefficient (CC)

	Nodes	Edges	Density	ASP	AC	CC
AMAZON	334863	925872	$8,25 * 10^{-06}$	2,78	-0,06	0,21

B. Overlapping community detection methods

In this section, we review the overlapping community detection methods used in this work. Community detection is a prolific subject in the literature, and a great variety of algorithms have been developed so far to deal with this issue. Some recent surveys have proposed taxonomies of the community-detection methods [13]. In this paper, we adopt the classification into five categories proposed by Xie et al.

[7]. These categories are Clique Percolation, Fuzzy Detection, Line Graph/Link Partitioning, Local Expansion/Optimization as well as Agent-Based/Dynamical Algorithms. Note that some algorithms do not belong to any of these categories. Table II reports some basic information about the algorithms considered. Note that, we selected these overlapping detection algorithms for one or more of the following reasons. They have been recently introduced, they are easily available on the web, they are popular in the literature.

TABLE II. ALGORITHMS USED FOR DETECTING THE OVERLAPPING COMMUNITY STRUCTURE. THE CLASSES ARE CLIQUE PERCOLATION (CP), LOCAL EXPANSION/OPTIMIZATION (LE/O), FUZZY DETECTION (FD), LINE GRAPH/LINK PARTITIONING (LG/LP), AND LABEL PROPAGATION (LP)

Algorithm	Classes	Reference	Complexity
CFINDER	CP	Palla et al. 2005 [14]	polynomial
LFM	LE/O	Lancichinetti et al. 2009 [8]	$O(n^2)$
MOSES	FD	McDaid et al. 2010 [15]	$O(en^2)$
GCE	LE/O	Lee et al. 2010 [16]	$O(mh)$
OSLOM	LE/O	Lancichinetti et al. 2011 [17]	$O(n^2)$
DEMON	LP	Coscia et al. 2012 [18]	$O(n+m)$
SLPA	LP	Xie et al.2012 [19]	$O(tm)$
SVINET	LG/LP	Prem et al.2013 [20]	not explicitly stated

In the following, we briefly review the operating principle of each class as well as the mechanism of the associated algorithms.

1) Clique Percolation: The main assumption of this approach is that a community is made of a combination of small network motifs called clique. In a network, a k-clique is a group of nodes of size k, such that every node is connected to each other node. In k-clique percolation, we say that two cliques of size k percolate into each other if they share $k-1$ nodes. A k-clique-community is the union of all k-cliques that can be reached from each other through a series of percolating k-cliques.

CFINDER[1], the Clique Finder, is one of the most popular overlapping community detection algorithms. It uses the Clique Percolation Method introduced by Palla et al. 2005 [14]. The algorithm first extracts all complete subgraphs of the network that are not part of a larger complete subgraph. The aim of the first phase is to populate a clique overlap matrix where its elements are equal to the number of common nodes between the corresponding two cliques. The diagonal entries are equal to the size of the clique. The k-clique-communities can be found by erasing every off-diagonal entry smaller than k. In the implementation of Adamcsek et al. 2005 [21] the parameter, k range from 3 to 8.

2) Fuzzy Detection: Fuzzy community detection algorithms build on the relation between links and communities. This association is made using a fitness function which differs from one method to another. The first step consists of randomly selecting links. Then, initial communities are formed which are composed of the extremity nodes of these links. After that, nodes are added to these communities by maximizing the function of fitness. Finally, the last step consists of successively removing the nodes of the communities to which they belong, and then see if the integration into another community would increase the fitness function.

MOSES[2], Model-Based Overlapping Seed ExpanSion, was proposed by McDaid et al. 2010 [15]. This method applies the Fuzzy Detection steps with a fitness function based on OSBM (Overlapping Stochastic Block Models) proposed by Latouche et al. 2011 [22]. The computational time complexity is equal to $O(en^2)$ where n is the number of nodes and e is the number of edges to be expanded.

3) Line Graph/Link Partitioning: If communities are defined as communities of nodes, we can also create communities of links. The basic idea of this approach is to define a projection graph in which the nodes represent the links of the original graph and the definition of a similarity value in order to understand how close two edges of the network are. A classical clustering algorithm can then be applied.

SVINET, the algorithm introduced by Gopalan and Blei [20], uses a Bayesian approach to detect overlapping communities. It assumes a probabilistic membership model of networks where each node can belong to multiple communities. Given an observed network, the model defines a posterior distribution that gives a decomposition of the nodes into overlapping communities. In particular, the posterior will place higher probability on configurations of the community memberships that describe densely connected communities. With this posterior, we can investigate which specific communities are responsible for each of the observed links. In this sense, the algorithm discovers link communities. As for many interesting Bayesian models, however, this posterior is intractable to compute. In order to approximate the posterior the algorithm iterates between subsampling the network, analyzing the subsample, and updating the estimated community structure. It is efficient because it only analyzes a subgraph of the network at each iteration.

4) Local Expansion/Optimization: Local Expansion/Optimization algorithms are based on growing a natural community. They generally perform in two steps. The first step is to find the initial communities called grains. These cores serve as seed communities for the second step of the process, that expands the cores by adding or removing nodes until a local density function cannot be improved. Non-overlapping algorithms can be used to find the initial grains. The second step is to contract these small grains in order to construct the final communities. The authors propose to add or remove some nodes to increase the community strength, defined as the ratio between internal and total node's degree of a community.

LFM[3], Lancichinetti Fortunato Method, expands a community from a random seed node by adding nodes until a fitness function is locally maximal. After finding one community, LFM randomly selects another node not yet assigned to any community to grow a new community. The fitness function control the strength of the community and the size of the communities:

[1]http://www.cfinder.org/

[2]https://sites.google.com/site/aaronmcdaid/moses

[3]https://github.com/sumnous/LFM_improve

$$f(c) = \frac{k_{in}^c}{(k_{in}^c + k_{out}^c)^\alpha} \quad (1)$$

where k_{in}^c and k_{out}^c are successively the internal and external nodes degree of the community c, and α is the resolution parameter controlling the size of the communities.

OSLOM[4], Order Statistics Local Optimization Method, is not based on a single idea. The authors propose to use Infomap or Louvain for detecting seed communities. This method does not detect all communities in very important recovery cases. In these situations, the authors have advocated the use of small grains communities created by taking a random node and by adding to it an arbitrary number of neighbors. Secondly, for each grain, OSLOM will apply rules to successively add and remove nodes until reaching a stable state in which it is no longer interesting to modify the community. The time complexity is $O(n^2)$, where n is the number of nodes.

GCE[5], Greedy Clique Expansion Lee et al. 2010 [16], is also on the same principle. The authors propose to use the maximal cliques as grains. Given the very large number of these initial grains, an optimization by a local fitness function is used to reduce this huge number. The time complexity for greedy expansion is $O(mh)$, where m is the number of edges, and h is the number of clique.

5) Label Propagation: Each node is initialized with a unique label. Then, each node replaces his label by the most figured label on its neighbors. In the case of equality, the label is randomly selected. After a number of iterations, nodes with the same label tends to be associated in communities. Therefore, all nodes having the same label form a community.

SLPA[6], Speaker-listener Label Propagation Algorithm, has been introduced by Xie et al. [19]. Each node has a labels memory. It updates its last label from the most frequent neighbor's memory label. SLPA has a time complexity equals to $O(tm)$ when m is the total number of edges and t is the memory size.

DEMON[7], introduced by Coscia et al. 2012 [18], tends to affect a node to the most frequent community given by the application of a label propagation algorithm on its neighbors sub-graph. The time complexity is equal to $O(n + m)$ where n is the number of nodes and m is the number of edges.

IV. Data analysis and discussion

In order to perform the analysis of the overlapping community structure, we build eight community networks. Recall that the nodes are the communities and the links between two nodes represent the fact that the communities overlap. For simplicity, we note the AMAZON community network based on the ground truth community structure as AMAZON*. As for the networks of communities built from the community structures extracted by each of the ten algorithms, we refer to them by the name of the algorithms from which they arise. Note that, in general, the overlapping community networks can have multiple components. We therefore applied a preprocessing step where all the small components are eliminated. The networks analysis is in any cases restricted to the largest weakly connected component.

We first present and compare classical basic global properties of these networks. We then turn the analysis of various distribution that summarize some important topological characteristics of these networks. Finally, we analyze the distribution more specific to the overlapping community structure.

A. Basic topological properties

Table III reports the global properties of the ground-truth based community network AMAZON* as well as the ones estimated through the various community detection algorithms. Looking at these results, the first comment that comes to mind is that there is a great variability of the behavior of the algorithms. All the algorithms under-estimate the number of overlapping communities. The algorithms can be classified in two groups according to the number of communities they detect. The first group is made of CFINDER, MOSES, SLPA, SVINET, DEMON. They find around 20000 communities as compared to around 75000 for AMAZON*. In the second group that includes LFM, GCE and OSLOM around 10000 communities are identified. The number of edges which represent the number of overlapping communities is much more smaller than in the reference. All the algorithms tend to underestimate the overlap between the communities. The network density values confirm this observation. Apart from SVINET, whose density is comparable to that of AMAZON*, it is one order of magnitude smaller for the other algorithms. The diameter values range from 16 to 37 while the reference is 27. Even if three out of height are smaller than the reference, we can say that at this level the networks are fairly consistent. According to the degree correlation estimated, there is a vast majority of assortative community networks while AMAZON* is disassortative. CFINDER is the only algorithm that lead to a disassortative network. Finally, all the networks are much more clustered than AMAZON*. To summarize this first set of results, the community networks originating from the community detection algorithms seems to be quite different than AMAZON*. These results are overall quite disappointing except for SVINET that seems to emerge from the crowd.

TABLE III. Global properties of AMAZON* and structural community networks. The calculated properties are number of nodes (Nodes), number of edges (Edges), Density, Diameter (D), Assortativity Coefficient (AC), and Clustering Coefficient (CC)

	Nodes	Edges	Density	D	AC	CC
AMAZON*	74698	1062092	3.8E-04	27	-0.16	0.02
CFINDER	21888	31522	6.5E-05	24	-0.02	0.15
LFM	8914	7585	9.5E-05	37	0.11	0.09
MOSES	25415	72499	1.1E-05	31	0.51	0.41
GCE	10256	13526	1.2E-05	31	0.25	0.13
OSLOM	9876	12613	1.2E-05	29	0.23	0.16
DEMON	17809	99293	3.1E-05	16	0.23	0.29
SLPA	25455	53442	8.2E-05	22	0.03	0.13
SVINET	25162	123947	3.9E-04	28	0.03	0.09

[4]http://oslom.org/

[5]https://sites.google.com/site/greedycliqueexpansion/

[6]https://sites.google.com/site/communitydetectionslpa/

[7]http://www.michelecoscia.com/?page_id=42

B. Degree distribution

Degree distribution measures the statistical repartition of the network nodes degrees. For a large number of networks, such distribution can be adequately described as a power-law that can be written as $(P(k) \sim k^{-\alpha})$, where α is a positive exponent. Related experimental studies show that the exponent value of the power law usually ranges from 2 to 3. Fig.1 reports the empirical degree distribution of the overlapping community networks together with their estimated distribution under the power-law hypothesis. The exponent value is computed using the maximum likelihood estimators described in [23]. From this viewpoint, results are much more satisfying. Indeed, the power-law seems to be a good fit for all the networks. Results of the Kolmogorov-Smirnov (KS) test reported in Table IV consolidate this intuition. Among the nine alternative distributions under test (Beta, Cauchy, Exponential, Gamma, Logistic, Log-Normal, Normal, Uniform and Weibull) the Log-Normal distribution is the only one which exhibits small KS values. The explanation may be that for low degree values, the empirical distribution is well approximated by the Log-Normal and that the Power-Law is a better fit for the tail. Note that this is not surprising as very similar basic generative models can lead to either Power-Law or Log-Normal distributions. Based on these results we can classify the algorithms in two groups. The first one include DEMON and SVINET that behave in the same way as AMAZON*. In this group, the Log-Normal hypothesis might be an alternative, while in the second group the Power-Law is the only serious hypothesis. When we look at the Power-Law exponent values reported in Table V, in any case it is always higher than the one estimated for AMAZON*. Nevertheless, it is worth noticing that they are in the range generally observed for most real-world complex networks. Other parameters such as Average node degree and Max node degree are quite disparate. It reflects the great variability of the networks basic properties.

TABLE IV. KS TEST VALUES FOR THE DEGREE DISTRIBUTION. THE DISTRIBUTION UNDER TEST ARE THE POWER-LAW (PL), BETA (BE), CAUCHY (CA), EXPONENTIAL (E), GAMMA (GM), LOGISTIC (LO), LOG-NORMAL (LN), NORMAL (N), UNIFORM (U), AND WEIBULL (WB)

	PL	BE	CA	E	GM	LO	LN	N	U	WB
AMAZON*	0.03	0.87	0.23	0.23	0.87	0.44	0.06	0.44	0.98	0.19
CFINDER	0.02	0.4	0.22	0.39	0.4	0.36	0.25	0.38	0.94	0.25
LFM	0.02	0.61	0.41	0.61	0.61	0.33	0.37	0.31	0.86	0.33
MOSES	0.03	0.23	0.23	0.23	0.23	0.25	0.13	0.27	0.83	0.22
GCE	0.02	0.45	0.25	0.45	0.45	0.27	0.24	0.28	0.84	0.27
OSLOM	0.03	0.41	0.29	0.41	0.41	0.25	0.24	0.24	0.82	0.29
DEMON	0.04	0.17	0.22	0.15	0.14	0.22	0.08	0.24	0.8	0.22
SLPA	0.01	0.3	0.24	0.3	0.3	0.29	0.17	0.31	0.9	0.26
SVINET	0.03	0.36	0.22	0.17	0.34	0.28	0.08	0.31	0.89	0.26

C. Average clustering coefficient as a function of degree

In order to estimate this distribution, we first compute the local clustering coefficient for every node in the network. Then, for each set of nodes that has the same degree, we compute the average clustering coefficient. For a large number of networks, this distribution can be adequately represented by a Power-Law [24]. Therefore, we draw the distribution in a logarithmic scale, as shown in Fig.2. Results of KS test values are reported in Table VI. In the light of these results, it is not easy to make a definitive conclusion. Indeed, according to the KS test values, there is three hypothesis that can be satisfactory for the distribution of AMAZON* (Power-Law, Log-Normal, Weibull). Nevertheless, it appears clearly on the plot that the Power-Law is a good fit for the tail. There is no dominant hypothesis for CFINDER, while the power-law is the most likely for LFM. For the remaining community networks, results are more mixed and multiple hypotheses can be satisfying. Indeed, there is between three and four hypothesis that can be a good fit if one consider that with a KS test value strictly smaller than 0.1, the hypothesis cannot be rejected. Note that the power-law is always one of them.

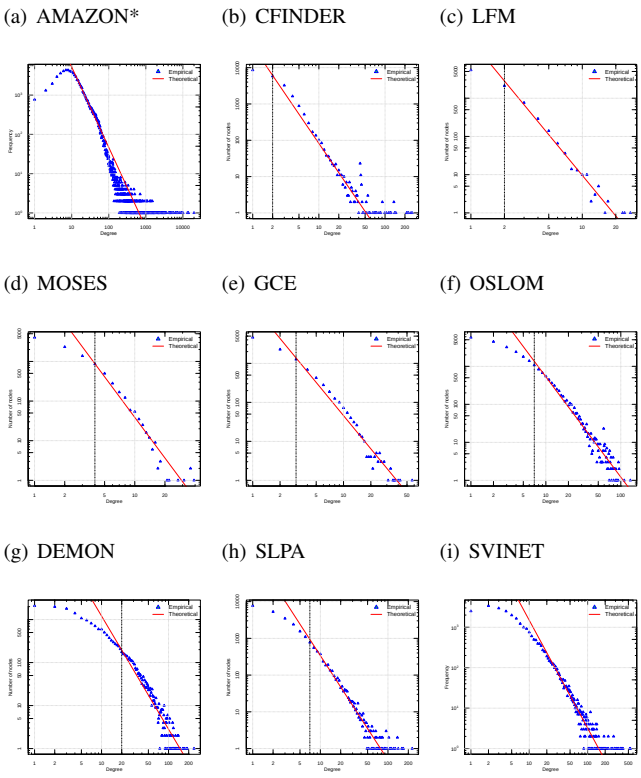


Fig. 1. Log-log empirical degree distribution (blue) and Power-Law estimating (red) of AMAZON* (a), CFINDER (b), LFM (c), MOSES(d), GCE (e), OSLOM (f), DEMON(g), SLPA (h) and SVINET (i)

TABLE V. GLOBAL PROPERTIES AND ESTIMATED DEGREE DISTRIBUTION PARAMETERS FOR AMAZON* AND DETECTED COMMUNITY NETWORKS. THE PARAMETER IS THE POWER-LAW EXPONENT(α)

	Average degree	Max degree	α
AMAZON*	28.43	19991	2.13
CFINDER	2.88	257	2.67
LFM	1.7	27	3.89
MOSES	5.71	134	3.14
GCE	2.64	57	3.51
OSLOM	2.55	39	3.79
DEMON	11.15	240	3.04
SLPA	4.2	228	3.02
SVINET	9.8	540	3.08

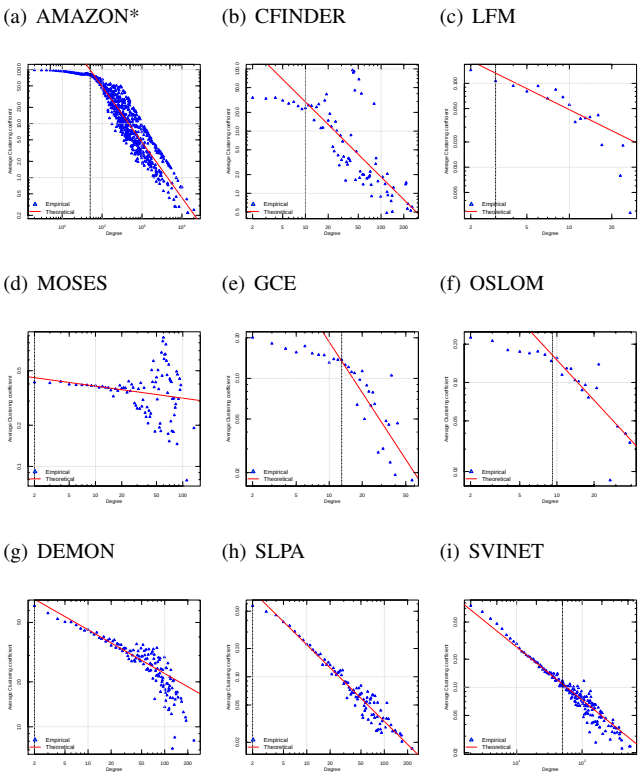


Fig. 2. Log-log empirical Average clustering coefficient distributions as a function of the degree (blue) and Power-Law estimating (red) of AMAZON* (a), CFINDER (b), LFM (c), MOSES(d), GCE (e), OSLOM (f), DEMON(g), SLPA (h), and SVINET (i)

TABLE VI. KS TEST VALUES FOR THE AVERAGE CLUSTERING COEFFICIENT AS A FUNCTION OF DEGREE. THE DISTRIBUTION UNDER TEST ARE THE POWER-LAW (PL), BETA (BE), CAUCHY (CA), EXPONENTIAL (E), GAMMA (GM), LOGISTIC (LO), LOG-NORMAL (LN), NORMAL (N), UNIFORM (U), AND WEIBULL (WB)

	PL	BE	CA	E	GM	LO	LN	N	U	WB
AMAZON*	0,03	0,39	0,21	0,1	0,37	0,31	0,04	0,33	0,93	0,05
CFINDER	0,16	0,19	0,24	0,2	0,2	0,13	0,21	0,14	0,73	0,31
LFM	0,07	0,15	0,19	0,15	0,15	0,15	0,1	0,14	0,47	0,2
MOSES	0,08	0,09	0,11	0,18	0,14	0,1	0,17	0,08	0,31	0,11
GCE	0,09	0,06	0,18	0,09	0,08	0,12	0,1	0,11	0,47	0,2
OSLOM	0,09	0,08	0,18	0,09	0,1	0,13	0,1	0,13	0,44	0,24
DEMON	0,06	0,04	0,16	0,1	0,08	0,1	0,12	0,09	0,43	0,16
SLPA	0,05	0,06	0,19	0,11	0,06	0,17	0,07	0,19	0,61	0,22
SVINET	0,05	0,07	0,23	0,08	0,04	0,17	0,09	0,19	0,65	0,04

D. Hop distance distribution

The hop plot represents the distribution of pairwise distances in a network. It is usually represented as a cumulative distribution. Fig.3 represents the cumulative distribution of the eight community networks. As shown in Table VII of KS test values, there is a clear evidence that the gaussian hypothesis is the best fit for all the networks except OSLOM. All the mean hop distance values are higher than the one measured for AMAZON*. This is also the case for the dispersion as measured by the standard deviation.

As shown in Table VII of KS test values, except OSLOM, the Gaussian distribution hypothesis outperforms all the other alternative hypotheses under test. SLPA and SVINET parameters of the Gaussian distribution are the nearest to the AMAZON* parameters.

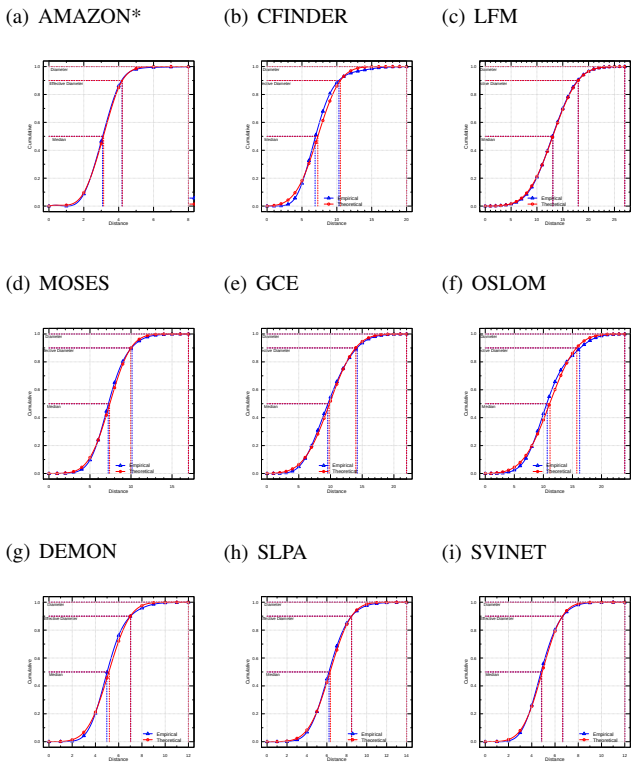


Fig. 3. Hop distance cumulative distributions for AMAZON* (a), CFINDER (b), LFM (c), MOSES(d), GCE (e), OSLOM (f), DEMON(g), SLPA (h), and SVINET (i)

TABLE VII. KS TEST VALUES FOR THE HOP DISTANCE. THE DISTRIBUTION UNDER TEST ARE THE POWER-LAW (PL), BETA (BE), CAUCHY (CA), EXPONENTIAL (E), GAMMA (GM), LOGISTIC (LO), LOG-NORMAL (LN), NORMAL (N), UNIFORM (U), AND WEIBULL (WB)

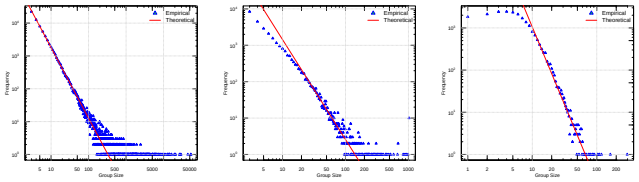
	PL	BE	CA	E	GM	LO	LN	N	U	WB
AMAZON*	0,4	0,27	0,59	0,66	0,22	0,41	0,43	0,05	0,86	0,91
CFINDER	0,26	0,27	0,1	0,31	0,34	0,29	0,51	0,03	0,18	0,48
LFM	0,13	0,31	0,22	0,66	0,25	0,8	0,26	0,05	0,29	0,61
MOSES	0,22	0,21	0,14	0,8	0,55	0,6	0,13	0,04	0,49	0,78
GCE	0,88	0,53	0,51	0,76	0,76	0,1	0,15	0,01	0,88	0,39
OSLOM	0,7	0,21	0,44	0,15	0,66	0,11	0,23	0,11	0,43	0,43
DEMON	0,43	0,41	0,74	0,8	0,19	0,46	0,63	0,01	0,09	0,82
SLPA	0,1	0,35	0,45	0,13	0,28	0,71	0,89	0,05	0,35	0,59
SVINET	0,75	0,8	0,73	0,87	0,61	0,45	0,67	0,06	0,72	0,29

E. Community size distribution

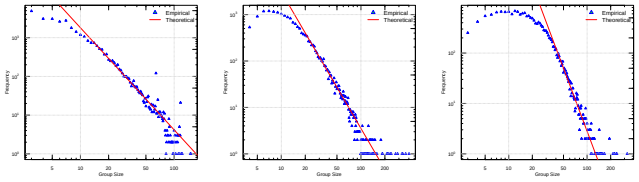
In [14], Palla et al. introduce four distributions in order to quantify the overlapping community structure in complex networks (the community degree, the community size, the membership number, the overlap size). The community degree distribution is just the degree distribution of the overlapping community network. The membership of a node is its number of communities and the size of a community is the number of nodes it contains. Previous analysis [14],[25] on real-world networks, have shown that these distributions (or at

least that the tail of these distributions) can be adequately described by a power-law. We performed a comparative study of the ground truth community structure of AMAZON and the community structures uncovered by the algorithms. Due to the lack of space, we report only results for the community size distribution and comment briefly on the other properties. We choose to report the community size distribution because it has been a widely studied property in real-world networks. Fig.4 reports the empirical distributions and the estimated Power-Law for the ground-truth community structure of AMAZON and the outputs of the community detection algorithm. The Power-Law seems to be a good fit in any case. This is confirmed by the results of the KS test reported in Table VIII. Indeed, the Power-Law exhibits the smallest KS value. Note that the Log-Normal is not far behind for most of the algorithms (LFM, MOSES, GCE, OSLOM, DEMON, SLPA and SVINET). Table IX reports the number of communities, the maximal community size, the average community size and the Power-Law exponent for the community structure under test. These results confirm our previous remarks about the great variability of the algorithms. Although, the community size distribution can be adequately described by a Power-Law, this common property results in a wide range of situations. From a qualitative point of view, the same comments can be made based on the results on the membership and overlap size distributions.

(a) AMAZON* (b) CFINDER (c) LFM



(d) MOSES (e) GCE (f) OSLOM



(g) DEMON (h) SLPA (i) SVINET

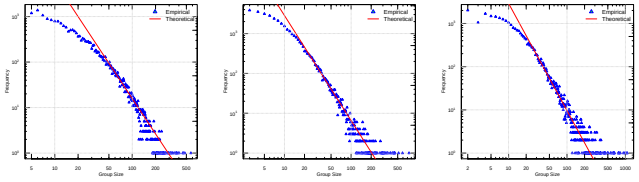


Fig. 4. Log-log empirical Community size distribution (blue) and Power-Law estimating (red) of AMAZON* (a), CFINDER (b), LFM (c), MOSES(d), GCE (e), OSLOM (f), DEMON(g), SLPA (h), and SVINET (i)

V. Conclusion

The main objective of this study is to highlight the need to focus on topological properties in order to evaluate community detection algorithms. Indeed, evaluation of community detection algorithms usually relies either on nodes classification or on quality metrics that encode a desirable community structure property. A good score is considered as the evidence that the uncovered community structures correspond to the underlying community structure. In this work, we present a comprehensive analysis of overlapping community structure of a large-scale real-world network. The ground-truth community structure is compared with the output of eight different overlapping community detection algorithms. To do so, we use the overlapping community networks where the nodes are the communities and the links describe the overlap between two communities. This allows us to analyze several properties of their topological structure. Results clearly show that no community detection algorithm is able to recover the ground truth community structure. Furthermore, there are substantial differences between the algorithms. But one must remain cautious because these results may be specific to the dataset. It is therefore crucial to extend this work by a systematic study of various typical real-world networks. Anyway, these results confirm the remarks reported in previous studies [1], [5], [12] that there are significant differences between ground truth communities and structural communities uncovered by community detection algorithms. Consequently, a more detailed analysis of the community structure is needed in order to validate the algorithms.

TABLE VIII. KS test values for the Community size. The distribution under test are the Power-Law (PL), Beta (BE), Cauchy (CA), Exponential (E), Gamma (GM), Logistic (LO), Log-Normal (LN), Normal (N), Uniform (U), and Weibull (WB)

	PL	BE	CA	E	GM	LO	LN	N	U	WB
AMAZON*	0,01	0.68	0.27	0.57	0.68	0.47	0.14	0.48	0.98	0.2
CFINDER	0.01	0.5	0.26	0.32	0.49	0.39	0.12	0.41	0.94	0.23
LFM	0.01	0.16	0.24	0.11	0.16	0.17	0.09	0.19	0.91	0.31
MOSES	0.03	0.19	0.24	0.16	0.16	0.23	0.07	0.25	0.78	0.2
GCE	0.02	0.19	0.21	0.06	0.18	0.19	0.05	0.22	0.86	0.27
OSLOM	0.02	0.07	0.22	0.1	0.06	0.13	0.07	0.13	0.81	0.27
DEMON	0.04	0.13	0.21	0.11	0.1	0.21	0.05	0.24	0.82	0.23
SLPA	0.02	0.3	0.23	0.14	0.29	0.28	0.07	0.3	0.91	0.25
SVINET	0.02	0.35	0.21	0.13	0.33	0.29	0.04	0.31	0.9	0.24

TABLE IX. The number of communities, communities maximum size, the communities average size and the Power-Law exponent for AMAZON and detected community structure

	Communities	Maximum size	Average size	α
AMAZON	75149	53551	30,23	2,08
CFINDER	28402	1023	10,16	2,55
LFM	21841	296	6,84	3,98
MOSES	30240	151	10,89	2,81
GCE	17043	402	16,32	4,09
OSLOM	17007	325	20,91	4,47
DEMON	19839	572	26,7	4,67
SLPA	33986	740	13,26	3,22
SVINET	25302	1073	19,51	2,86

References

[1] J. Yang and J. Leskovec, "Structure and overlaps of ground-truth communities in networks." *ACM TIST*, vol. 5, no. 2, p. 26, 2014.

[2] W. M. Rand, "Objective Criteria for the Evaluation of Clustering Methods," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, Dec. 1971.

[3] A. N. Albatineh, M. Niewiadomska-Bugaj, and D. Mihalko, "On Similarity Indices and Correction for Chance Agreement," *Journal of Classification*, vol. 23, no. 2, pp. 301–313–313, Sep. 2006.

[4] M. Warrens, "On the equivalence of cohens kappa and the hubert-arabie adjusted rand index," *Journal of Classification*, vol. 25, no. 2, pp. 177–183, 2008.

[5] G. K. Orman, V. Labatut, and H. Cherifi, "Comparative evaluation of community detection algorithms: a topological approach," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, no. 08, p. P08001, 2012.

[6] B. Abrahao, S. Soundarajan, J. Hopcroft, and R. Kleinberg, "A separability framework for analyzing community structure," *ACM Trans. Knowl. Discov. Data*, vol. 8, no. 1, pp. 5:1–5:29, Feb. 2014.

[7] J. Xie, S. Kelley, and B. K. Szymanski, "Overlapping community detection in networks: The state-of-the-art and comparative study." *ACM Comput. Surv.*, vol. 45, no. 4, p. 43, 2013.

[8] A. Lancichinetti, S. Fortunato, and J. Kertesz, "Detecting the overlapping and hierarchical community structure of complex networks," 2008.

[9] D. W. Collins and W. T. Dent, "A comparison of alternative testing methodologies used in capital market research," *Journal of Accounting Research*, vol. 22, no. 1, pp. 48–84, 1984.

[10] S. Harenberg, G. Bello, L. Gjeltema, S. Ranshous, J. Harlalka, R. Seay, K. Padmanabhan, and N. Samatova, "Community detection in large-scale networks: a survey and empirical evaluation," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 6, no. 6, pp. 426–439, 2014.

[11] J. Yang and J. Leskovec, "Defining and evaluating network communities based on ground-truth." *Knowl. Inf. Syst.*, vol. 42, no. 1, pp. 181–213, 2015.

[12] D. Hric, R. K. Darst, and S. Fortunato, "Community detection in networks: Structural communities versus ground truth," *Phys. Rev. E*, vol. 90, p. 062805, Dec 2014.

[13] M. Coscia, F. Giannotti, and D. Pedreschi, "A classification for community discovery methods in complex networks," *Statistical Analysis and Data Mining*, vol. 4, no. 5, pp. 512–546, 2011.

[14] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society." *Nature*, vol. 435, no. 7043, pp. 814–8, 2005.

[15] A. F. McDaid and N. J. Hurley, "Detecting highly overlapping communities with model-based overlapping seed expansion." in *ASONAM*. IEEE Computer Society, 2010, pp. 112–119.

[16] C. Lee, F. Reid, A. McDaid, and N. Hurley, "Detecting highly overlapping community structure by greedy clique expansion," in *Workshop on Social Network Mining and Analysis*, 2010.

[17] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato, "Finding statistically significant communities in networks," *PLoS ONE*, vol. 6, no. 4, p. e18961, 04 2011.

[18] M. Coscia, G. Rossetti, F. Giannotti, and D. Pedreschi, "Demon: a local-first discovery method for overlapping communities." in *KDD*, Q. Y. 0001, D. Agarwal, and J. Pei, Eds. ACM, 2012, pp. 615–623.

[19] J. Xie, B. K. Szymanski, and X. Liu, "Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process." in *ICDM Workshops*. IEEE Computer Society, 2011, pp. 344–349.

[20] P. K. Gopalan and D. M. Blei, "Efficient discovery of overlapping communities in massive networks," *Proceedings of the National Academy of Sciences*, vol. 110, no. 36, pp. 14 534–14 539, 2013.

[21] B. Adamcsek, G. Palla, I. J. Farkas, I. Dernyi, and T. Vicsek, "Cfinder: locating cliques and overlapping modules in biological networks." *Bioinformatics*, vol. 22, no. 8, pp. 1021–1023, 2006.

[22] P. Latouche, E. Birmelé, and C. Ambroise, "Overlapping stochastic block models with application to the French political blogosphere," *The Annals of Applied Statistics*, vol. 5, no. 1, pp. 309–336, Mar. 2011.

[23] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Rev.*, vol. 51, no. 4, pp. 661–703, Nov. 2009.

[24] X.-Q. Cheng, F.-X. Ren, S. Zhou, and M.-B. Hu, "Triangular clustering in document networks," *New Journal of Physics*, vol. 11, no. 3, p. 033019, Mar. 2009.

[25] M. Jebabli, H. Cherifi, C. Cherifi, and A. Hammouda, "Overlapping community structure in co-authorship networks: A case study," in *u- and e- Service, Science and Technology (UNESST)*, 2014, pp. 26–29.