Multidimensional networks

A novel node centrality metric based on common neighborhood

Gamgne Domgue Félicité, Tsopze Norbert, Ndoundam René

Sorbonne University, IRD, UMMISCO, F-93143, Bondy, France,
University of Yaounde I
Yaounde,Cameroon
felice.gamgne@gmail.com, felicite.gamgne@uy1.uninet.cm
tsopze@uy1.uninet.cm
ndoundam@gmail.com

ABSTRACT. Complex networks have been receiving increasing attention by the scientific community. They can be represented by multidimensional networks in which there is multiple types of connections between nodes. Thanks also to the increasing availability of analytical measures that have been extended in order to describe and analyze properties of entities involved in this kind of multiple relationship representation of networks. These measures focused on quantitative involvement of node through the widely popular and intuitive measure of degree. However, one aspect of such properties have been disregarded so far: entities in such networks are often tied according to the interest they have to their neighbors in the overall dimensions. In this paper, the problem of characterizing multidimensional networks, using a *qualitative* aspect of the node neighborhood, has been studied, through the new defined node centrality measure, *Stability*, to describe the connectivity of nodes that incorporates across-dimension topological features in order to identify the relevant dimensions. We assessed our measure on two real-world multidimensional networks, showing its validity, its meaningfulness and its correlation with a dimension connectivity measure.

RÉSUMÉ. Les réseaux complexes ont reçu beaucoup d'attention de la part de la recherche scientifique. Ils peuvent etre représentés par des réseaux multidimensionnels dans lesquels il existe plusieurs types de relations entre les entités. Plusieurs propriétés decrivant les noeuds et permettant l'extraction de la connaissance sur de tels réseaux, ont été étudiées. La majorité d'entre elles prone l'aspect quantitatif de la connectivité d'un noeud, à l'instar de la centralité de degré. Cependant, un aspect primordial a été omis: celui *qualitatif*, basé sur le type de voisinage d'un noeud. En effet, dans de tels réseaux, les entités sont généralement connectées selon les mêmes centres d'intérêts qu'ils possèdent. Dans ce travail, le problème de caractérisation des réseaux multidimensionnels moyennant l'utilisation de la notion qualitative du voisinage d'un noeud, a été abordé à travers la définition d'une nouvelle mesure de centralité appelée *Stabilité*. Cette dernière permet de décrire la connectivité des noeuds basée sur les caractéristiques topologiques, en vu de déterminer les dimensions pertinentes. L'évaluation de cette mesure s'effectue sur deux réseaux multidimensionnels réels, et montre sa validité et sa correlation à une mesure de connectivité de dimensions.

KEYWORDS : Multidimensionnal networks, Centrality measure, Relevance dimension

MOTS-CLÉS : Reseaux multidimensionnels, mesure de centralité, Pertinence de dimension

1. Introduction

Complex network analysis has received a lot of attention by the scientific researchers, because it helps to better understand the intrinsic behavior of relationships between entities. These relationships could be either of one or several types. Unlike a *monodimensional* network which contains only one type of links between nodes, multidimensional networks contain links which either reflect different kinds of relationship or represent different values of the same kind of relationship among a same set of elementary components. This flexibility allowed to use complex networks to study real-world systems in many fields: sociology, physics, genetics, computer, etc. Such systems can be modeled by multidimensional networks as reported in Figure 1 [1] where on the left we have different types of links, while on the right we have different values (conferences) for one relationship (for example, co-authorship). Multidisciplinary and extensive research works have been devoted to the extraction of non trivial knowledge from such networks [7]. Some of them focused on the characterization of their properties. More precisely, they studied some centrality measurements based on the quantitative neighborhood also called "weight", of each node [1, 6]. These measures, which are certainly relevant, do not take into account a more recent reality. Indeed, with the advent of the Internet and social networking sites, individuals communicate more easily when the majority of their contacts use the same platforms or means of communication as they do. Thus, the qualitative aspect of this neighborhood is important to be considered. From this aspect, a semantics emerges relating to the retention of the same neighbors of a node over all dimensions, namely *Stability*. To the best of our knowledge, however, the literature still misses a systematic qualitative measure for weight-based centrality in the context of correlated multidimensional networks, together with a model of extracting relevant dimensions for each node. The aim of this paper is precisely defining a basic and analytical concept of centrality measure, which takes into account the connectivity redundancy of nodes among dimensions. As questioned in [1], how is it possible to contribute to answering the question *To what extent one or more dimensions are more important than others for the connectivity of a node?*

Contributions: In this work:

– We introduce a novel centrality weighting scheme of nodes called *stability*, in multidimensional network

– We formally define a measure aimed at extracting useful knowledge on relevance dimensions of nodes

– We characterize nodes of the multidimensional network according to their stability

– we empirically test the meaningfulness of our measure, by means of a case study on two realistic networks.

The rest of the paper is organized as follows. Section 2 overviews related works, Section 3 describes the proposed measures to assess the activity level of a node in a dimension. Section 4 presents experimental evaluation, Section 5 concludes the paper.

2. Related works

Multidimensional networks have for a long time been proposed as an alternative to better describe interactions within complex systems [1]. For instance, in social networks, individuals can be connected according to different social ties, such as friendship or family relationship [2]. The extraction of knowledge and analysis of both the local and global properties of such networks remains of interest to scientists. Indeed, multidimensional networks abound with a large amount of information, particularly concerning the various kinds of relationship between entities. Since an individual may have a particular interest for a certain number of dimensions, he could be influential or important in regard to other nodes in these dimensions: they are then qualified as central. So ignoring centrality in multidimensional structures can lead to different ranking results than what one obtains for multidimensional networks [8] .

Centrality, an indicator that quantifies the importance of nodes in a network, comes from the discipline of Social network analysis and has become a fundamental concept in network science with its applications in a range of disciplines. In recent works, many efforts have been devoted to "centrality" measures in order that they are also applicable in multidimensional networks [8]. Examples of these various centrality measures include degree centrality, called overlapping degree in [6]. As they help to extract a knowledge and analyze the network properties related to the questioning of “how important a dimension for a node is”, these measures are based on both relevance dimension and dimension connectivity, since nodes could exist across all dimensions. A multigraph used to model a multidimensional network is denoted by a triple $G = (V, E, L)$ where: V is a set of nodes; L is a set of dimensions; E is a set of edges, i.e the set of triples (u, v, d) where $u, v \in V$ are nodes and $d \in L$ is a dimension. Thus, Berlingerio [1] defined a relevance dimension measure based on the connectivity of dimensions, as described in Equation 1, which computes the fraction of neighbors directly reachable from node v following edges belonging only to the set of dimensions called D with $D \subseteq L$. Likewise, he defined a measure Node Exclusive Dimension Connectivity ($NEDC$) computing the ratio of nodes belonging only to a specific dimension d, as described in Equation 2.

$$DR_{XOR}(v, D) = \frac{|Neighbors_{XOR}(v, D)|}{|Neighbors(v, L)|} \tag{1}$$

$$NEDC(d) = \frac{|u \in V | \exists v \in V : (u, v, d) \in E \land \forall j \in L, j \neq d : (u, v, j) \notin E|}{|u \in V | \exists v \in V : (u, v, d) \in E|} \tag{2}$$

where $Neighbors_{XOR}(v, D)$ is the set of neighbors of v belonging only to dimensions D. Despite their popularity and effectiveness in social network analysis, we believe to our knowledge that these measures mainly take into account the quantitative aspect (i.e. degree) of node properties and links, yet the *qualitative* aspect, namely the type of neighboring nodes, would have a significant impact on facilitating communication between an individual and his neighborhood.

Indeed, these measurements focus only on the degree of nodes regardless of the type of neighborhood, to extract the relevant dimensions. According to them, a relevant dimension for a node is quantified by the density of its neighborhood. Thus, if a node has the same number of neighbors on all dimensions, then all these dimensions will be relevant to it. However, in real life situations relating to human relationships, the communication is more obvious, more easy or cheaper among individuals using the same platform of information exchanges. So, the interest a user has for a platform depends on the subscription of his friends to that platform. Therefore, a dimension(platform) would be more relevant for a node(subscriber) if the node has a conservative behavior of its neighborhood over all dimensions. It is this concept that we implement in the next section.

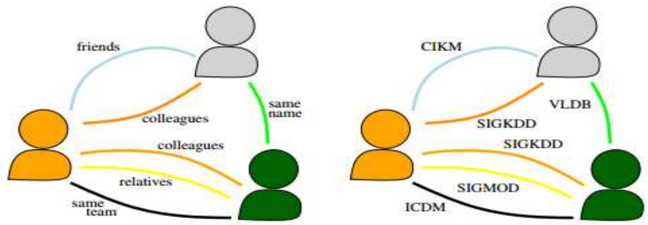


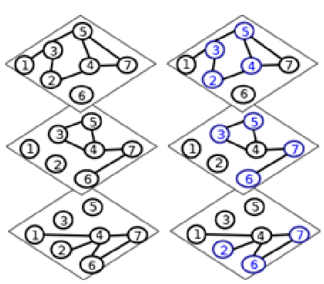
Figure 1: Example of multidimensional networks

3. Multidimensional network analysis

This paragraph presents new defined measures to contribute to knowledge extraction from a multidimensional networks. They concern whether a dimension can be of interest for a node, based on the stability of his neighbors. In the first subsection, the stability centrality is defined, and in the second subsection, we present how to determine the relevant dimensions of a node.

3.1. Stability centrality

The weights of the nodes have been the subject of some studies. According to graph theory, this weight corresponds to the sum of the weights of the edges incident to this node. It is a variant of the degree centrality. This measure shows that the importance of a node depends on the number of communications it establishes with its neighborhood. It corresponds to its activity level in a network. Extending this measure in multidimensional networks, Nicosia et al. [6] studied that the activity of a node in a particular dimension is very often correlated with its activity in another dimension. The authors considered the centrality degree as a measure of the node activity in a dimension. However, the number of neighbors seems to be meaningless when studying behavior of entities in a context of correlated dimensions. Then it becomes necessary to maintain the stable behavior of a node in order to make easy information exchange among its community membership. The stability centrality of a node u is then pointed up and computes the proportion of the common neighborhood of this node between two dimensions p and q, through a *Jaccard index* similarity as defined through Definition 1. It takes into account the structural features across several dimensions.



	1	2	3	4	5	6	7
Dim1	0.0	0.25	0.125	0.45	0.33	0.0	0.33
Dim2	0.0	0.0	0.125	0.33	0.33	0.25	0.66
Dim3	0.0	0.25	0.0	0.28	0.0	0.25	0.66

	1	3	2	6	5	4	7
ε	0.0	0.083	0.166	0.166	0.22	0.35	0.55

Figure 2: An example of connected multidimensional network with 3 dimensions on the left, and on the right, the stability node on the top and threshold for relevance dimension extraction below.

Definition 1 ***(The stability centrality of a node in a dimension)***. *Stability centrality of node u in dimension q measures the common neighborhood of a node between q and the other dimensions. The function $Stability : V \times D \to [0,1]$ is defined as:*

$$Stability(u,q) = \frac{1}{ndim-1} \sum_{p=0}^{ndim-1} \frac{\mid \Gamma_u^p \cap \Gamma_u^q \mid}{\mid \Gamma_u^p \cup \Gamma_u^q \mid} \tag{3}$$

whereΓ_u^p denotes the neighborhood of node u in the dimension p and $ndim$ denotes the number of dimensions. We refer to *disassortative stability* when its neighborhood is totally different in all dimensions; Stability tends to be null. Otherwise, it is the *assortative stability*; it tends to its maximal value 1. In this paper, the node with the lowest disassortative stability is unstable and the one with the highest assortative stability is the most stable over the network. As shown in table of Figure 2 above, node 1 possesses a disassortative stability, unlike node 7 which gets an assortative stability.

3.2. Relevance dimension

The concept of dimension relevance of a node studied in [5] stresses on that dimension in which the node has the most important exclusive degree as defined by Berlingerio [1] i.e. it computes the fraction of neighbors directly reachable from node u following edges belonging exclusively to a subset of dimensions D_l as shown in Equation 1. This way does not seem relevant in some real situations, because if a node has the same degree on all dimensions of the network, then all of them will be relevant. Yet, if we consider only those in which the node has a more stable neighborhood, the relevance of the dimensions would be more semantic. The relevant dimensions $RD(u)$ of a node u refers to those dimensions for which the node has a stability centrality greater than or equal to a certain threshold ε. It is described by the function $RD : V \to D$ as:

$$RD(u) = \{q, | Stability(u,q) \geq \varepsilon\} \tag{4}$$

The threshold ε is defined in the Equation 5. When the node u has a stability centrality whose value is higher than ε, it is said that the node u *is stable for the subset of dimensions* $RD(u)$, or that the *dimensions in the subset $RD(u)$ are relevant for the node u.*

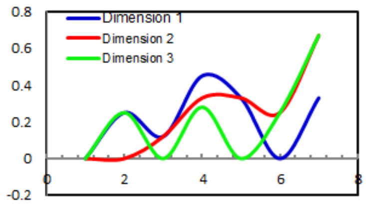
$$\varepsilon = \frac{1}{|D|} \sum_{i=1}^{|D|} Stability(u,i) \tag{5}$$

4. Experimental Evaluation

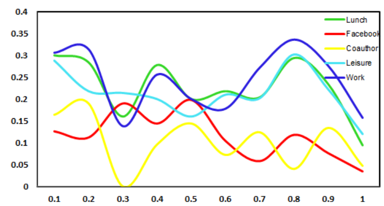
In this section, we assess the proposed metric on two main sights: its correlation with dimension connectivity and the behavior of nodes according to the values of the metric.

4.1. Correlation with dimension connectivity

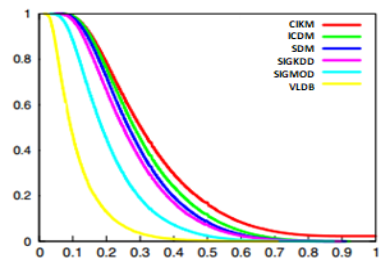
This section reports the results obtained by computing the stability measure on two real-world multidimensional network datasets, namely AUCS [4] and DBLP [3]. AUCS, an attributed multidimensional network, models relationships between 61 employees of Aarhus University Computer Science department considering five different aspects: coworking, having lunch together, Facebook friendship, offline friendship, and coauthorship. In



(a) Stability centrality on the network in Fig. 2 illustration



(b) Cumulative Stability on AUCS



(c) Cumulative Stability on DBLP

Figure 3: Stability centrality distribution

DBLP, there are 83901 nodes which correspond to authors, tied by 159302 links, and 50 dimensions represent the top-50 Computer Science conferences. Two authors are connected on a dimension if they co-authored at least two papers together in a particular conference. All the experiments were conducted on an Intel Core $i5-8250U$ CPU $@1.60GH_z$, $8GB$ of RAM machine, Windows 10 OS 64 bytes.

Figure 3 reports the cumulative distribution of the stability measure. It denotes the average of nodes' stability on 10 intervals. The latter corresponds to the normalization of the number of nodes. Figure 3(a) is a small dataset. Then there is no need to cumulate the stability centrality of the nodes, unlike the figures 3(b) and 3(c) whose size is important, leading to a normalization of their x-axis. Berlingerio in [1] analyzed the correlation between the DR_{xor} distribution and the Dimension Connectivity values (especially $NEDC$). The authors deduced that DR_{xor} measure is correlated to the $NEDC$ measure. Following them, we analyze the correlation between the stability distribution and the $NEDC$ measure. What can be seen by looking at the Stability distribution and $NEDC$ values, reported in Tables 1-3, is that the Stability distributions seem to be correlated to the $NEDC$ measure. This correlation is not surprising since by definition, the two measures are two different perspectives, one local (Node stability) and one global (Dimension Connectivity), of the same aspect: how much a node is important for the connectivity of a network. We note, in fact, that the stability tends to be higher in conjunction with higher $NEDC$ values.

Table 1: Node connectivity, Node stability, computed on the illustration network in Fig. 2

Dimension	Stability average	NEDC
Dimension 1	0.63	0.85
Dimension 2	0.4	0.7
Dimension 3	0.47	0.7

Table 2: Node connectivity, Node stability, computed on AUCS network

Dimension	Stability average	NEDC
Lunch	0.22	0.15
Facebook	0.10	0.03
Coauthor	0.11	0.05
Leisure	0.20	0.11
Work	0.24	0.25

4.2. Analyzing node behavior

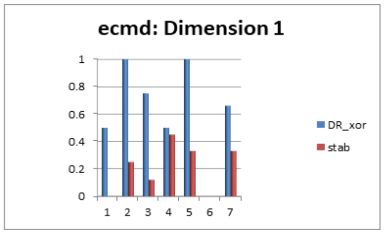
This section describes node behaviors in the overall dimensions, according to its degree and its stability. Assume that the network in Figure 2 represents an exchange of experiences through a multidimensional network between actors of the agriculture area (e.g. farmers), in which a dimension describes a type of crop (banana, onion, rice, etc.). The idea behind the stability metric is that a stable node/actor is the more important because it favors a success in agricultural business, moreover, the other nodes/actors trust in him. Then, the stability of a farmer's neighborhood demonstrates his competence; therefore, he becomes a more reliable source of information. An individual may not be reliable if he loses his regular relationship. In Figure 2, node 7 is an example of this. If the second and third dimensions are removed, it looses its trusted contacts, but still remains present in the network. On the other hand, according to DR_{xor} measure, that node 7 disappears from the network. As shown in Figure 4, this node has the same value of DR_{xor} across all dimensions, but its value of stability is low in the first dimension. Otherwise, the node 1 has a disassortative behavior. Any dimension is relevant for this node, meaning that it is less important across dimensions.

5. Conclusion

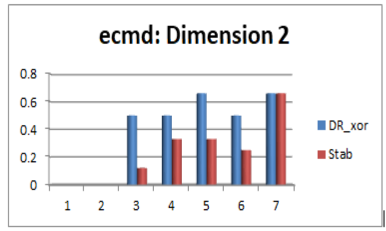
We proposed a novel centrality measure based on stability of the neighborhood of nodes. Since an active node on one dimension can remain inactive on the rest of the dimensions, it is possible to study the stability of a node in a multidimensional context, according to its center of interest. Therefore we defined the notion of relevance dimension of a node in order to contribute to the question of how important is a dimension for a node. An assessment on the semantic given to the stability centrality compared to degree centrality was carried out. Likewise, we show how correlated the stability to the dimension connectivity NEDC measure. The next study intends to assess the impact of this measure on the communities obtained from a well-defined model.

Table 3: Node connectivity, Node stability, computed on DBLP network

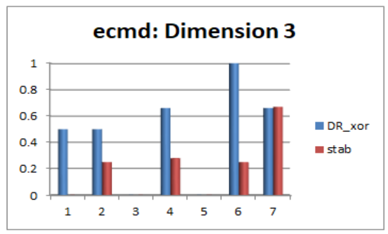
Dimension	Stability average	NEDC
VLDB	0.00115	0.75
SIGMOD	0.0046	0.97
CIKM	0.15	3.86
SIGKDD	0.0087	1.38
ICDM	0.098	2.45
SDM	0.055	1.44



(a) Stability behavior on dimension 1



(b) Stability behavior on dimension 2



(c) Stability behavior on dimension 3

Figure 4: Stability and DR_{xor} assessments on network in Fig. 2

6. References

[1] M. Berlingerio, M. Coscia , F. Giannotti , A. Monreale , M. Coscia , D. Pedreschi, “Multidimensional networks: foundations of structural analysis”, *World Wide Web*, pp 567–593, (2013).

[2] M. Verbrugge Lois, “Multiplexity in Adult Friendships*”, *Social Forces* vol. 57 (4), pp 1286-1309, (1979).

[3] B. Boden , S. Günnemann , H. Hoffmann , T. Seidl, “Mining coherent subgraphs in multi-layer graphs with edge labels”, *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* pp 1258–1266, (2012).

[4] M. Magnani , B. Micenkova , R. Rossi, “Combinatorial analysis of multiple networks”, *rXiv preprint arXiv:1303.4986*, (2013).

[5] O. Boutemine , M. Bouguessa, “Mining community structures in multidimensional networks”, *ACM Transactions on Knowledge Discovery from Data (TKDD)* vol. 11,(4), pp 51, (2017).

[6] V. Nicosia , V. Latora, “Measuring and modeling correlations in multiplex networks”, *Physical Review E*, vol. 92, (3), pp 032805, (2015).

[7] P. Mucha , T. Richardson , K. Macon , J. Onnela, “Community structure in time-dependent, multiscale, and multiplex networks”, *American Association for the Advancement of Science*, vol. 328, (5980), pp 876–878, (2010).

[8] M. De Domenico , A. Sole-Ribalta , E. Omodei , S. Gòmez , A. Arenas, “Centrality in interconnected multilayer networks”, *American Association for the Advancement of Science*, vol. , (CoRR abs/1311.2906), (2013).