On multidimensional network measures*

Matteo Magnani, Anna Monreale, Giulio Rossetti, and Fosca Giannotti

KDDLab, ISTI, CNR, via G. Moruzzi 1, 56124 Pisa, Italy
name.surname@isti.cnr.it

Abstract. Networks, i.e., sets of interconnected entities, are ubiquitous, spanning disciplines as diverse as sociology, biology and computer science. The recent availability of large amounts of network data has thus provided a unique opportunity to develop models and analysis tools applicable to a wide range of scenarios. However, real-world phenomena are often more complex than existing graph data models. One relevant example concerns the numerous types of social relationships (or edges) that can be present between individuals in a social network. In this short paper we present a unified model and a set of measures recently developed to represent and analyze network data with multiple types of edges.

Keywords: Multidimensional networks, multi-layer networks, centrality measures

1 Introduction

Recent years have witnessed a dramatic increase of interest in the modeling power provided by network models, representing the relationships between abstract entities. This interest has been determined by the wide applicability of these models to several disciplines, including social network analysis and statistics, biology, physics and economics [**?**,**?**,**?**], and by the contemporary availability of large network datasets, a primary example being online social networks.

The interdisciplinary character of this problem has influenced recent developments in the field of computer science, where simple graph models have been enriched with additional information to conform to the representation needs required in other disciplines. For example, in the context of social network data one important feature to be considered is the differentiation between several kinds of connections: the relationship between two individuals is a complex phenomenon that cannot be simplified by reducing it to a simple edge between two nodes in a graph. *Working together*, *being friend* or *having a connection on LinkedIn* are very different kinds of ties. They all contribute to the definition of a complex social relationship, but cannot be merged and downgraded to a mono-dimensional concept.

This consideration has been described decades ago in social sciences [**?**] and later formalized in the field of social network analysis [**?**], but only recently it has

* This work has been partly supported by FIRB project RBFR107725.

been incorporated in data models aimed at managing and mining large social databases [?,?,?]. In this short paper we describe recent efforts of the authors to formalize this concept and define new analysis tools.

In particular, we focus on centrality measures. These are single-node or whole-network functions that characterize the structure of the network and the role of its nodes inside it. Centrality measures constitute the primary analysis tool for simple networks, and in this work we present their extension to the context of multidimensional networks. Specifically, the contribution of this work is the first unified presentation of the main measures for multidimensional networks proposed by the authors in recent years, re-defined on a common data model.

The paper is organized as follows. In the next section we introduce our unified data model and extended centrality measures. We then show how these measures can be used to better understand the role of different kinds of connections inside a complex network. We conclude the paper with a discussion of our view over this research area and our priorities for future directions.

2 A model for multidimensional networks

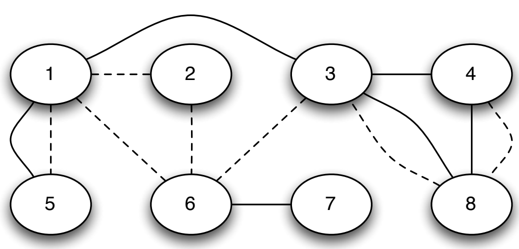

Fig. 1. Example of a multidimensional network

In literature, many analytical measures, both at the local and at the global levels, have been defined in order to describe and analyze properties of standard, monodimensional networks. However networks are often multidimensional: therefore, multidimensional analysis is needed to distinguish among different kinds of interactions or, equivalently, to look at interactions from different perspectives. Analytical measures come under a different light when seen in this setting, since the analysis scenario gets even richer, thanks to the availability of different dimensions to take into account. As a consequence, in this novel setting it becomes indispensable: (a) to define a model able to fully represent multidimensional networks and their properties; (b) to study how most of the measures defined for classical monodimensional networks can be generalized in order to be applied to multidimensional networks; (c) to define new measures, meaningful only in the multidimensional scenario, to capture hidden relationships among different dimensions.

We choose to use a *multigraph* to model a multidimensional network. For the sake of simplicity, in our model we only consider undirected multigraphs and since we do not consider node labels, hereafter we use *edge-labeled undirected multigraphs*, denoted by a triple $\mathcal{G} = (V, E, L)$ where: V is a set of nodes; L is a set of labels; E is a set of labeled edges, i.e. the set of triples (u, v, d) where $u, v \in V$ are nodes and $d \in L$ is a label. Also, we use the term *dimension* to indicate *label*, and we say that a node *belongs to* or *appears in* a given dimension d if there is at least one edge labeled with d adjacent to it. We also say that an edge *belongs to* or *appears in* a dimension d if its label is d. We assume that given a pair of nodes $u, v \in V$ and a label $d \in L$ only one edge (u, v, d) may exist. Thus, each pair of nodes in $\mathcal{G}$ can be connected by at most $|L|$ possible edges. Hereafter $\mathcal{P}(L)$ denotes the power set of L. An example of multidimensional network is shown in figure 1.

3 Multidimensional network measures

Given a multidimensional network we can extend typical measures defined on traditional monodimensional networks and we can define new measures that are meaningful only in this specific setting.

In general, in order to adapt classical measures to the multidimensional setting we need to extend the domain of each function in order to specify the set of dimensions for which they are calculated. Intuitively, when a measure considers a specific set of dimensions, a filter is applied on the multigraph to produce a view of it considering only that specific set, then the measure is calculated over this view.

3.1 Degree

In order to cope with the multidimensional setting, we can define the degree of a node w.r.t a single dimension or a set of them. To this end, we have to redefine the domain of the classical degree function by including also the dimensions.

Definition 1 (Degree). *Let $v \in V$ be a node of a network G. The function $Degree : V \times \mathcal{P}(L) \rightarrow \mathbb{N}$ defined as*

$$Degree(v, D) = |\{(u, v, d) \in E \ s.t. \ u \in V \ \wedge \ d \in D\}|$$

computes the number of edges, labeled with one of the dimensions in D, between v and any other node u. □

We can consider two particular cases: when $D = L$ we have the degree of the node v within the whole network, while when the set of dimensions D contains only one dimension d we have the degree of v in the dimension d, which is the classical degree of a node in a monodimensional network.

As an example, node 3 in Figure 1 has degree centrality $Degree(3, \{d_1, d_2\}) = 5$, indicating its five adjacent edges, and if we focus only on the dashed dimension this reduces to 2.

3.2 Neighborhood

In classical graph theory the *degree* of a node refers to the connections of a node in a network: it is defined, in fact, as the number of edges adjacent to a node. In a simple graph, each edge is the sole connection to an adjacent node. In multidimensional networks the degree of a node and the number of nodes adjacent to it are no longer related, since there may be more than one edge between any two nodes. For instance, in Figure 1, node 3 has four neighbors and degree equal to 5 (taking into account all the dimensions). In order to capture this difference, we define the following:

Definition 2 (Neighbors). *Let $v \in V$ and $D \subseteq L$ be a node and a set of dimensions of a network $G = (V, E, L)$, respectively. The function $Neighbors : V \times \mathcal{P}(L) \to \mathbb{N}$ is defined as*

$$Neighbors(v, D) = |NeighborSet(v, D)|$$

where $NeighborSet(v, D) = \{u \in V \mid \exists (u, v, d) \in E \ \wedge \ d \in D\}$. This function computes the number of all the nodes directly reachable from node v by edges labeled with dimensions belonging to D. □

Note that in the monodimensional case the value of this measure corresponds to the degree. It is easy to see that $Neighbors(v, D) \leq Degree(v)$, but we can also easily say something about the ratio $\frac{Neighbors(v,D)}{Degree(v)}$. When the number of neighbors is small, but each one is connected by many edges to v, we have low values of this ratio, which means that the set of dimensions is somehow redundant w.r.t. the connectivity of that node.

We also define a variant of the Neighbors function, which takes into account only the adjacent nodes that are connected by edges belonging only to a given set of dimensions.

Definition 3 (Neighbors$_{XOR}$). *Let $v \in V$ and $D \subseteq L$ be a node and a set of dimensions of a network $G = (V, E, L)$, respectively. The function $Neighbors_{XOR} : V \times \mathcal{P}(L) \to \mathbb{N}$ is defined as*

$$Neighbors_{XOR}(v, D) = |\{u \in V | \ \exists d \in D : (u, v, d) \in E \wedge \nexists d' \notin D : (u, v, d') \in E\}|$$

As an example, in we consider the dashed dimension in Figure 1 the only XOR-neighbor of node 3 is node 6. This indicates how this dimension is fundamental in keeping the two nodes connected — which is not the case for, e.g., nodes 3 and 8, that would be anyway connected through the other dimension.

3.3 Multidimensional distance

The main assumption underlying the definition of multidimensional distance is that different edge types are incomparable to each other: if Renzo and Lucia are married and Agnese is Lucia's mother, who is *closer* to Lucia? Our answer is to consider the two relationships as alternative ways of being connected that should not be reduced to a monodimensional concept when distances are computed [**?**].

For example, in Figure 1 we can go from node 3 to node 4 by traversing the continuous edge (3 —— 4), or through two steps along the dashed edge (3 - - - 8 - - - 4) or through a combination of the two (3 —— 8 - - - 4). The number of steps taken in each network determine the multidimensional length of these paths.

Definition 4 (Multi-layer path length). *The multi-layer path length of path* p *is an array* $r_1 + \ldots + r_{|L|}$ *where* r_i *indicates the number of edges traversed in the* i^{th} *dimension.* □

We may also include network switches in this definition – however this better applies to a model emphasizing the separation between the different networks [**?**], so in this paper we stick to this simpler definition.

The interesting aspect of multidimensional paths is that the choice of not comparing edges of different kinds does not prevent us to identify *shortest* paths containing different edge types. The situation where one multidimensional path is considered shorter than another is formalized using the concept of multidimensional *path dominance*.

Definition 5 (Path dominance). *Let* r *and* s *be two multi-layer path lengths.* r *dominates* s *iff* $\forall l \in [1, |L|]\ r_l \leq s_l \wedge \exists i\ r_l < s_l$. □

As an example, (3 —— 4) and (3 - - - 8 - - - 4) are both shortest paths between nodes 3 and 4, while (3 —— 8 - - - 4) is not: (3 —— 4) is shorter (or *dominates* it) because it involves the same number of steps in the continuous dimension and less steps in the other.

We can thus define the distance between two nodes n_1 and n_2 as follows:

Definition 6 (Multidimensional distance). *Let* $ML(n_1, n_2)$ *be set of all multidimensional path lengths between nodes* n_1 *and* n_2*. The distance between* n_1 *and* n_2 *is a set* $P \subseteq ML$ *such that* $\forall p \in P\ \nexists p' \in ML : p'$ *dominates* p. □

This definition has some attractive properties: it returns all paths that can be shortest under some monotone path evaluation function (e.g., a function assigning a weight to each dimension), does not return any path that cannot be the shortest given some evaluation function, and reduces to traditional path length in the monodimensional case. So, this extension is *tight* (sound and complete) and conservative.

3.4 Multidimensional betweenness centrality

Using the concept of multidimensional shortest path introduced in the previous section we can easily compute the betweenness centrality of any node, which characterizes the importance of the node with respect to blocking or favouring information propagation in the network.

In single networks, betweenness centrality is defined as the number of shortest paths passing through a given node. The same definition can be used to characterize betweenness in a multidimensional network by replacing the concept of shortest path with the one introduced in the previous section [**?**].

While we cannot provide an in-depth analysis of this measure for space reasons, it is worth noticing how multidimensional betweenness can capture the role of some nodes that would not emerge in a monodimensional setting. As an example, consider node 8 in Figure 1. If we compute its betweenness centrality in the *flattened* network where we do not distinguish between different kinds of edges we can see that no shortest path passes through it, leading to a centrality of 0. On the contrary, the betweenness centrality computed using our measure is positive. For example, a shortest multidimensional path between node 3 and node 4 might pass through 8 in the case where the dotted dimension is significantly stronger than the continuos one. This comes from the fact that the multidimensional distance between nodes 3 and 4 contains the two alternative paths (3 —— 4) and (3 - - - 8 - - - 4).

4 Dimension Relevance

One key aspect of multidimensional network analysis is to understand how important a particular dimension is over the others for the connectivity of a node. In the following we introduce the concept of *Dimension Relevance*, which for space reasons is only applied to the *neighborhood* measure — the same approach can be envisioned for path-based measures.

Definition 7 (Dimension Relevance). *Let $v \in V$ and $D \subseteq L$ be a node and a set of dimensions of a network $G = (V, E, L)$, respectively. The function $DR : V \times \mathcal{P}(L) \rightarrow [0, 1]$ is defined as*

$$DR(v, D) = \frac{Neighbors(v,D)}{Neighbors(v,L)}$$

and computes the ratio between the neighbors of a node v connected by edges belonging to a specific set of dimensions in D and the total number of its neighbors. □

Clearly, the set D might also contain only a single dimension d, for which the analyst might want to study the specific role within the network.

However, in a multidimensional setting, this measure may still not cover important information about the connectivity of a node. As an example, this measure does not capture the fact that for a node a dimension could be the only one that allows reaching a subset of its neighbors. To capture this aspect we introduce a variant of this measure.

Definition 8 (Dimension Relevance XOR). *Let $v \in V$ and $D \subseteq L$ be a node and a set of dimensions of a network $G = (V, E, L)$, respectively. DR_{XOR} : $V \times \mathcal{P}(L) \rightarrow [0, 1]$ is defined as*

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

and computes the fraction of neighbors directly reachable from node v following edges belonging only to dimensions D. □

In the following, we want to capture the intuitive intermediate value, i.e. the number of neighbors reachable through a dimension, weighted by the number of alternative connections.

Definition 9 (Weighted Dimension Relevance). *Let $v \in V$ and $d \in L$ be a node and a dimension of a network $G = (V, E, L)$, respectively. The function $DR_W : V \times \mathcal{P}(L) \to [0, 1]$, called* Weighted Dimension Relevance*, is defined as*

$$DR_W(v,D) = \frac{\sum_{u \in NeighborSet(v,D)} \frac{n_{uvd}}{n_{uv}}}{Neighbors(v,L)}$$

where: n_{uvd} is the number of dimensions which label the edges between two nodes u and v and that belong to D; n_{uv} is the number of dimensions which label the edges between two nodes u and v. □

The Weighted Dimension Relevance takes into account both the situations modeled by the previous two definitions. Low values of DR_W for a set of dimensions D are typical of nodes that have a large number of alternative dimensions through which they can reach their neighbors. High values, on the other hand, mean that there are fewer alternatives.

5 Discussion

In this work we have described a unified model and set of measures to represent and analyze multidimensional networks, i.e., networks with multiple edge types. Some of these measures have already been shown to be useful to identify structural features of a network, e.g. the role of specific dimensions [?], and to study network's hubs [?]. However, they constitute only a first foundational step towards a general extension of complex network mining methods to this multidimensional framework.

Some initial steps have already been taken in this direction, proposing preliminary extensions of well known graph mining problems like link prediction [?] and community detection [?,?]. A work that can benefit from the proposed model as well is [?] where the authors study the evolution of networks defining a way to identify Eras and turning points. However, we envision the need for a combination of theoretical advancements and collection of new datasets to corroborate the proposed methods. In fact, the ultimate objective is to be able to identify hidden and emerging patterns in data coming from real applications, so that these analytical approaches can be adopted by researchers and analysts from different disciplines.

In addition to these more practically-oriented objectives, the model described in this work provides a general reasoning framework to develop a better understanding of the social dynamics underlying the observed reality and shaping the collected data. Initial efforts in this direction have been made through preliminary studies concerning the dynamics of network formation [**?**] and the characterization of the so-called *tie strength* in multidimensional networks [**?**]. In our opinion, further development in these directions will require a stronger collaboration between disciplines and in particular a joint modelling effort between computer scientists and domain experts, to tune models that are both computationally tractable and provide an accurate representation of the reality under study.

Acknowledgements. This research work has been partially supported by European FP7 project DATASIM (http://www.datasim-fp7.eu/).

References

1. Wasserman, S., Faust, K.: Social Network Analysis. Cambridge University Press (1994)
2. Newman, M.E.J.: Networks: an introduction. Oxford University Press (2010)
3. Easley, D., Kleinberg, J.: Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press (2010)
4. Goffman, E.: Frame analysis: an essay on the organization of experience. Harper colophon books ; CN 372. Harper & Row, New York (1974)
5. Magnani, M., Rossi, L.: The ml-model for multi layer network analysis. In: IEEE International conference on Advances in Social Network Analysis and Mining, IEEE Computer Society, Los Alamitos (2011) isbn: 978-0-769-54375-8.
6. Berlingerio, M., Coscia, M., Giannotti, F., Monreale, A., Pedreschi, D.: Multidimensional networks: foundations of structural analysis. World Wide Web Journal (2012) 1–27
7. Brodka, P., Stawiak, P., Kazienko, P.: Shortest Path Discovery in the Multi-layered Social Network. In: 2011 International Conference on Advances in Social Networks Analysis and Mining, IEEE (July 2011) 497–501
8. Magnani, M., Rossi, L.: Pareto distance for multi-layer network analysis. In: International Conference on Social Computing, Behavioral Modeling and Prediction, Springer, Berlin (2013)
9. Magnani, M., Micenková, B., Rossi, L.: Combinatorial analysis of multiple networks (2013)
10. Berlingerio, M., Coscia, M., Giannotti, F., Monreale, A., Pedreschi, D.: The pursuit of hubbiness: Analysis of hubs in large multidimensional networks. Journal of Computer Science **2**(3) (2011) 223–237
11. Rossetti, G., Berlingerio, M., Giannotti, F.: Scalable link prediction on multidimensional networks. In: ICDM Workshops. (2011) 979–986
12. Berlingerio, M., Coscia, M., Giannotti, F.: Finding and characterizing communities in multidimensional networks. In: ASONAM. (2011) 490–494
13. Berlingerio, M., Coscia, M., Giannotti, F., Monreale, A., Pedreschi, D.: Evolving networks: Eras and turning points. Intelligent Data Analysis **17**(1) (2013) 27–48
14. Magnani, M., Rossi, L.: Formation of multiple networks. In: International Conference on Social Computing, Behavioral Modeling and Prediction, Springer, Berlin (2013) Accepted.
15. Pappalardo, L., Rossetti, G., Pedreschi, D.: "how well do we know each other?" detecting tie strength in multidimensional social networks. In: ASONAM. (2012) 1040–1045