Technical Report: Analysis of Hubs in Large Multidimensional Networks

Michele Berlingerio* Michele Coscia† Fosca Giannotti‡
Anna Monreale § Dino Pedreschi¶

November 2, 2009

Abstract

Hubs in complex networks are important nodes in terms of their connectivity to the whole network. In a mono-dimensional network, i.e., where only one kind of interaction is possible among nodes, the concept of hub has been widely studied, and it is at the basis of many important applications such as web search and epidemic outbreaks. However, in real world scenarios, networks are multidimensional, i.e., several possible kinds of connections exist among the nodes. In this setting, the concept of a hub should take into account the multiple dimensions, that can have varying influence on the connectivity of each node, and whose interplay can be relevant to assess the importance of an entity. In this paper, we tackle the problem of analyzing the relevance of dimensions for node connectivity, and how this relevance analysis can highlight hubs with peculiar, interesting behaviors in a large network. To this end, we consider the multidimensional generalization of the degree, namely the number of neighbors of a node, and a newly introduced class of measures, that we call *Dimension Relevance*. We show how to efficiently compute these simple measures on one of the possible representations of a multidimensional network, the multigraph. Moreover, we illustrate the usage of our new measures on two different real world networks: a word-word graph built on a search engine query log, and a popular large online social network, Flickr. In both cases, our proposed measures allow us to discover hubs for which one specific dimension is of high relevance and ensures a high connectivity of that node within the network. We advocate that the presented methodology covers a wide range of possible applications, from search engines to computer networks, from biological to social networks, where the interplay among different dimensions can really make the difference for the behavior of specific important entities.

*ISTI-CNR, Pisa, Italy
†University of Pisa, Italy
‡ISTI-CNR, Pisa, Italy
§University of Pisa, Italy
¶University of Pisa, Italy

1 Introduction

In this techincal report we present the proof of a theorem omitted in our work "Analysis of Hubs in Large Multidimensional Networks", submitted for publication to WWW2010.

2 Details on the Multidimensional Network Measures Section

Theorem 1 *Let $v \in V$ and $D \subseteq L$ be a node and a set of dimension of a multidimensional network $G = (V, E, L)$, respectively. It holds:*

$$DimRelevance_{XOR}(v, D) \leq DimRelevance_{W}(v, D) \leq$$

$$\leq DimRelevance(v, D).$$

Proof. In order to prove this theorem it is sufficient to show that

$$Neighbors_{XOR}(v, D) \leq \sum_{u \in NeighborSet(v,D)} \frac{n_{uvd}}{n_{uv}} \quad (1)$$

and

$$\sum_{u \in NeighborSet(v,D)} \frac{n_{uvd}}{n_{uv}} \leq Neighbor(v, D) \quad (2)$$

as $DimRelevance_{XOR}(v, D)$, $DimRelevance_{W}(v, D)$ and $DimRelevance(v, D)$ have the same denominator. Let:

$$A = Neighbors_{XOR}(v, D)$$
$$B = \sum_{u \in NeighborSet(v,D)} \frac{n_{uvd}}{n_{uv}}$$
$$C = Neighbors(v, D).$$

First of all, we prove the inequality (1). If the node v is connected to a neighbor u only by edges labeled with dimensions in D then in both A and B, u contributes with 1; if they are connected only by edges labeled with dimensions that do not belong to D then in both the formulas, A and B, u contributes with 0; finally, if they are connected by some edges labeled with dimensions in D and some edges labeled with dimensions that do not belong to D then in A the node u contributes with a value equal to 0 while in B it contributes with a value greater than 0. So, we have that $A \leq B$.

Now, we prove the inequality (2). If the node v is connected to a neighbor u only labeled with dimensions in D then in both the formula B and C it contributes with 1; if they are connected only by edges labeled with dimensions that do not belong to D then in A and B u contributes with 0; finally, if they are connected by some edges labeled with dimensions that do not belong to D and some edges labeled with dimensions in D then in B the node u contributes with a value equal to $\frac{n_{uvd}}{n_{uv}} < 1$ $(d \in D)$ while in C it contributes with 1. So, we have that $B \leq C$.