Influence and Passivity for Multilayer Networks

Ganga Reddy Tankasala
The Ohio State University
tankasala.1@osu.edu

Rashmi Rao Jayathirtha
The Ohio State University
rao.307@osu.edu

Harihara Varma Indukuri
The Ohio State University
indukuri.3@osu.edu

Srinivasan Parthasarathy
The Ohio State University
parthasarathy.2@osu.edu

1. INTRODUCTION

In most of the network scenarios we represent an engineering or a natural system by taking into account only one type of relationship or interaction between the entities of the network. This kind of interactions give rise to single layer networks. For example in a Twitter network the interaction between nodes are either a tweet, re-tweet or a reply. But in reality most of the engineering systems consist of entities that interact with each other in complicated patterns that can encompass multiple types of relationships, change in time, and include other types of complications. Such systems include multiple subsystems and layers of connectivity, and it is important to take such "multilayer" features into account to try to improve our understanding of complex systems. Consequently, it is necessary to generalize "traditional" network theory by developing (and validating) a framework and associated tools to study multilayer systems in a comprehensive fashion. The origins of such efforts date back several decades and arose in multiple disciplines, and now the study of multilayer networks has become one of the most important directions in network science.

In this project we came up with one such framework that analyses a node's or entity's relative influence and passivity in a single layer as well as a multilayer network setup. Influence and Passivity scores were first introduced in [1]. In [1] the authors came up with a algorithm that calculates IP for a single layer. Influence of a node is defined as the amount of influence it can have on other nodes.The influence measure is also dependent on active/passivity of nodes it influences and Passivity of a node is a measure of how difficult it is for other nodes to influence it.An active/participating node can be influenced easily compared to a passive/consuming node.

We devise a general model for influence using the concept of passivity for a generalised network and develop an efficient algorithm similar to the HITS algorithm to quantify the influence of all the users in the network. We calculate IP scores for multilayer networks using the same concept of IP for single layer networks. We calculate IP scores for synthetic networks as well as real networks. Datasets considered for the experiments included DBLP dataset.

The remaining of the report is organised as follows: In section 2 we cover the problem statement, section 3 describes the methodology of the project, in section 4 we explain in detail our experimental results, in section 5 we compare our algorithm with HITS algorithm, in section 6 we give important conclusions of our work and finally in section 7 we talk about the future work to be done.

Keywords

Multilayer, Twitter, network theory, influence, passivity, HITS

2. PROBLEM STATEMENT

There are a number of state of art algorithms like HITS, Pagerank to find the influence and passivity of a node in a network. But all of them take into account only the number of edges connected to the node and the information about the edge weight is not leveraged. In this project we aim to come up with an algorithm to find these measures by taking into account physically qualitative edges to find the influence and passivity. Apart from measuring the node's IP in a single layer, the same algorithm is extended to a multilayer network with similar nodes to estimate the influence and passivity of any given node by taking in to consideration its behaviour in multiple layers.

3. METHODOLOGY

Influence and Passivity scores as described in [1] for a single layer network are calculated using:

$$I_i \leftarrow \sum_{j:(i,j)\in E} u_{ij} P_j \tag{1}$$

where I_i is influence of i^{th} node

$$P_i \leftarrow \sum_{j:(j,i)\in E} v_{ij} I_j \tag{2}$$

where P_i is influence of i^{th} node

$$u_{ij} = \frac{w_{i,j}}{\sum\limits_{k:(k,j)\in E} w_{kj}} \tag{3}$$

$$v_{ji} = \frac{(1 - w_{j,i})}{\sum\limits_{k:(k,j)\in E} (1 - w_{jk})} \tag{4}$$

where w_{ij} is weight of edge from node i to node j and u_{ij}, v_{ji} are acceptance rate and rejectance rate of a node.

We represent u_{ij} and v_{ji} as A (Acceptance) and R (Rejectance) respectively. The closed form solution for IP is given by

$$I^k = \frac{\overrightarrow{A} * (R^T A)^{k-1} * \overrightarrow{P}(0)}{||\overrightarrow{A} * (R^T A)^{k-1} * \overrightarrow{P}(0)||_1} \tag{5}$$

$$I^k = \frac{(R^T A)^{k-1} * \overrightarrow{P}(0)}{||(R^T A)^{k-1} * \overrightarrow{P}(0)||_1} \tag{6}$$

Influence scores are given by dominant eigen vector of AR^T.

3.1 Multilayer IP

The main idea of this project was to calculate IP values for a multilayer network using the generalised equations of a single layer network. Multilayer networks usually consist of 2 or more layers interconnected to each other. So to calculate IP for a multilayer networks we take into consideration A and R for individual layers. We define Acceptance and Rejectance values separately for intra-layer connections and inter-layer connections. For intra-layer connections we name A, R as A^δ (A Delta) and R^δ (R Delta) and for inter-layer connections A^* (A Star) and R^* (R Star).

Now the effective A and R for multilayer network is given as:

$$A = A^\delta + A^*$$

$$R = R^\delta + R^*$$

3.2 A and R for Intra-layers

To calculate A^δ and R^δ we take into account acceptance and rejectance of a node in each layer. So let us take $A(i, j)$ and $R(i, j)$ as measure of acceptance and rejectance from node i to node j respectively. Therefore, A^δ and R^δ can be written as

$$A^\delta = (A^1 + A^2 + ... + A^n)$$

$$R^\delta = (R^1 + R^2 + ... + R^n)$$

where $A^1, A^2.....A^n$ and $R^1, R^2.....R^n$ are the acceptance and rejectance in layer1, layer2...layerN.

After computing A^δ and R^δ we modify them to be column, row Stochastic respectively.

3.3 A and R for Inter-layers

Similar to intra-layer A and R, we calculate A^* and R^* by taking into account the flows from node i in one layer to node j in another layer. Note that the inter-layer edges are directed. So the effective A^* and R^* are given by

$$A^* = \sum A^*(i, j)$$

$$R^* = \sum R^*(i, j)$$

where

$$A^*(i, j) = I_i * A(inter - layer) * I_j$$

$$R^*(i, j) = I_i * R(inter - layer) * I_j$$

where I_i, I_j are identity matrix of i^{th} and j^{th} layer and A (inter-layer) and R (inter-layer) refers to inter layer acceptance and rejectance values respectively.

We modify A^* and R^* to be column and row stochastic respectively such that

$$\sum A^*(\forall k, j) = 1$$

$$\sum R^*(i, \forall k) = 1$$

Finally the dominant eigen vector of $(A^\delta + A^*) * (R^\delta + R^*)^T$ will result in influence scores.

4. EXPERIMENTAL RESULTS

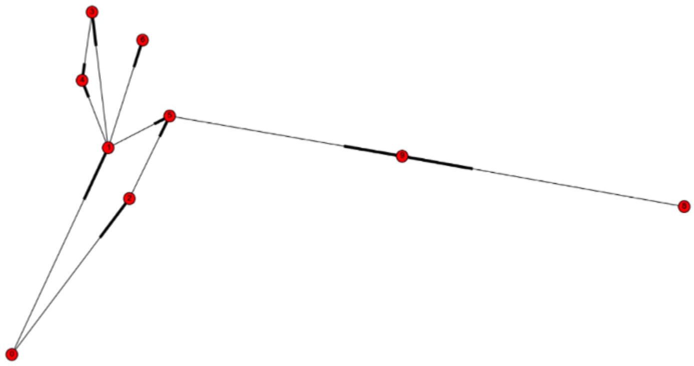
The algorithm was experimented on Single layer Synthetic Network, Multilayer Synthetic Network and also on DBLP dataset. The experimental results led to insights that provide a proof of the ability of our IP algorithm to give the Influence and Passivity values of nodes over the entire network.

4.1 Single layer Synthetic Network

We generated a synthetic single flow network of 9 nodes by assigning random weights to the edges. HITS and our IP algorithm were run on this network and the results are reported in the Table 1. Authority and Hub scores are obtained from HITS algorithm and Influence and Passivity scores are obtained form IP algorithm. The Hub score and the Authority score correspond to the Influence and Passivity factors respectively. From the table, it can be observed that the nodes 4 and 5 have similar Authority and Hub Scores while the edge weight factor is reflected in the difference in the Influence and Passivity values.

Node	0	1	2	3	4	5	6	8	9
Auth	0.0	0.01	0.01	0.43	0.56	0.56	0.43	0.0	0.02
Hub	0.93	0.01	0.26	0.26	0.0	0.00	0.00e	0.0	0.00e
Passivity	0.0	0.09	0.07	0.08	0.12	0.23	0.06	0.0	0.33
Influence	0.29	0.31	0.18	0.09	0.0	0.08	0.0	0.06	0.00

((a)) Table 1



((b)) Single layer Synthetic Network

Figure 1

4.2 Multilayer Synthetic Network

We generated 2 Synthetic networks to observe the correlation of inter connectivity of the layers and their IP values as well as the results from the HITS algorithm. (Fig.4) shows the IP values , Authority and Hub scores for the first synthetic network and (Fig.5) the corresponding values for second synthetic network. The color gradient is used to indicate the influence and passivity of the nodes. Blue color indicates higher score while yellow indicates lower score. The observations we made are as follows:

- Any node with high influence score in a layer has a low passivity score.
- A node with more inter-layer connections has higher influence over the entire network.
- When there are very few inter-layer connections the IP scores have a good correlation with hub and authority scores.

4.3 Multilayer network of DBLP Dataset

DBLP Dataset was chosen to construct a real multilayer network. We chose to represent the layers as the journals, nodes in each layer as authors and an edge between authors who have collaborated with each other in 1 or more papers. The weight of the edge has been assigned as the fraction of the total out degree of the node. Similar edge weight assignment is done with the inter layer nodes. Influence and Passivity were calculated for each of the nodes in the network.

The journals chosen are "Computer Graphics and Image Processing", "Computer Vision, Graphics, and Image Processing", "CVGIP: Graphical Model and Image Processing". The journals were chosen in related fields so as to get a common set of authors. (Fig.2b) shows the IP values obtained for all the authors and (Fig.2a) the 3D representation of the multilayer network. The IP scores of each layer are plotted in (Fig.3 and Fig.4).

The observations we made are as follows:

- Authors Azriel Rosenfeld, Larry S. Davis and Robert M. Haralickas have high Influence values.
- All the above authors with high influence values were present among the very few authors who contributed to all the three chosen journals.
- The Influence values of these authors on each layer reveals that a higher influential Author in one layer may not be more influential over the entire network of Journals and is aptly captured by our algorithm.
- Inter-layer connections have a greater impact on the influence and passivity scores.

5. RELATION TO STATE OF ART ALGORITHMS

When we compare IP with HITS algorithm, one of the main difference between IP and HITS is that, IP measure utilizes both the structural properties of the network as well as the diffusion behavior among users. The influence of a user thus depends not only on the size of the influenced audience, but also on their passivity. This differentiates IP from HITS, which was primarily based on individual statistical properties.

In HITS we compute authority and hub scores which correspond to influence and passivity in IP. The authority and hub score are calculated as follows:

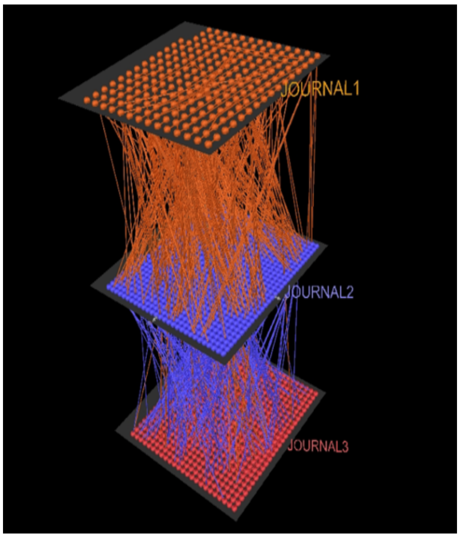
$$a_i^{t+1} \leftarrow \sum_{j:(j\rightarrow i)} h_j^{(t)} \tag{7}$$

where a_i^{t+1} is authority score of i^{th} node

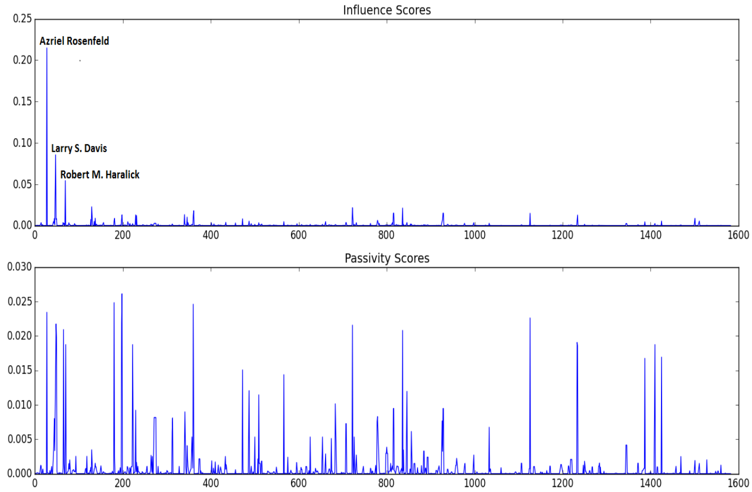
$$h_i^{t+1} \leftarrow \sum_{j:(i\rightarrow j)} a_j^{(t+1)} \tag{8}$$

where h_i^{t+1} is hub score of i^{th} node

We can see that authority score of a node depends only on the hub score of other nodes and vice versa.

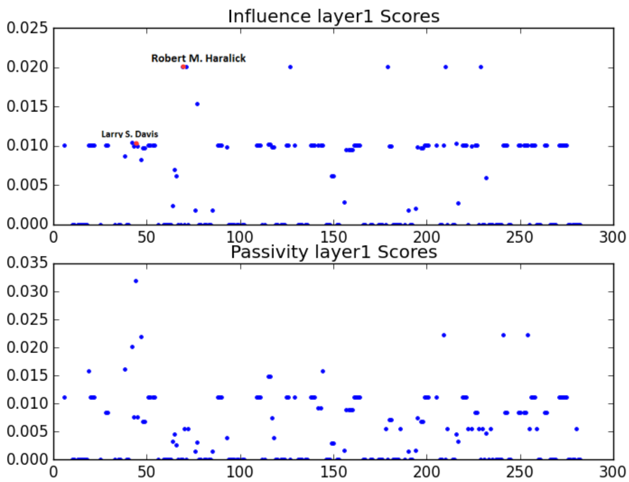


((a)) 3D representation of DBLP dataset for 3 journals

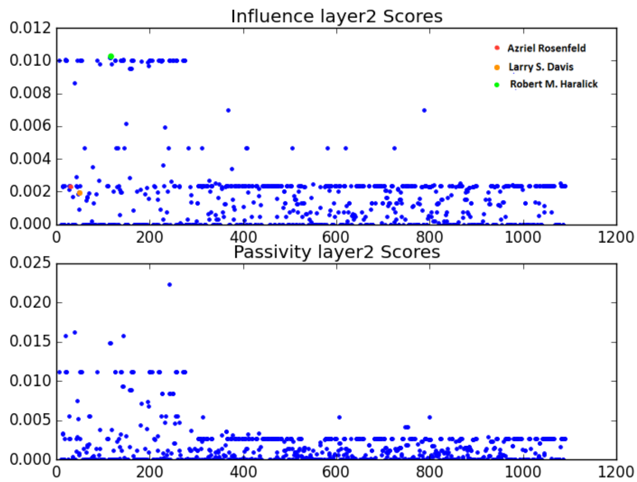


((b)) IP for DBLP DataSet

Figure 2

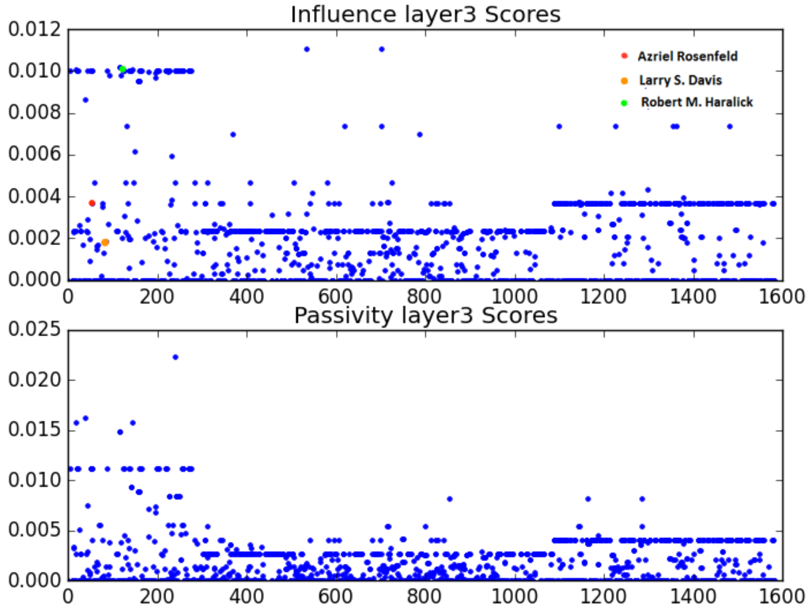


((a)) IP values for Journal1



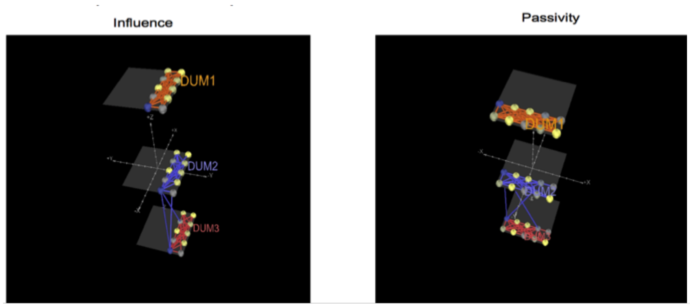
((b)) IP values for Journal2

Figure 3

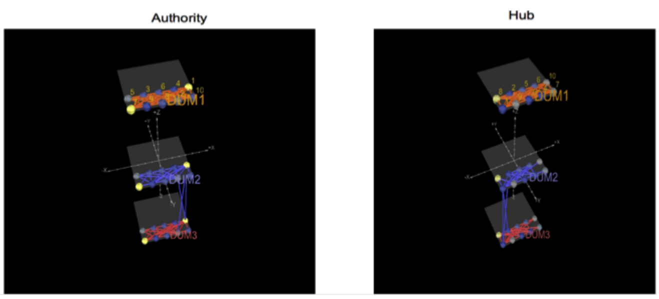


((a)) IP values for Journal3

Figure 4

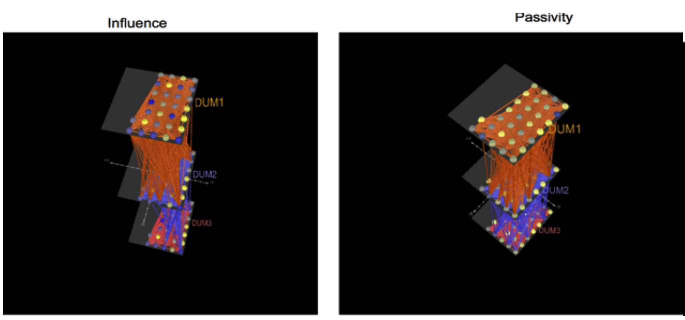


((a)) Synthetic Network 1 IP

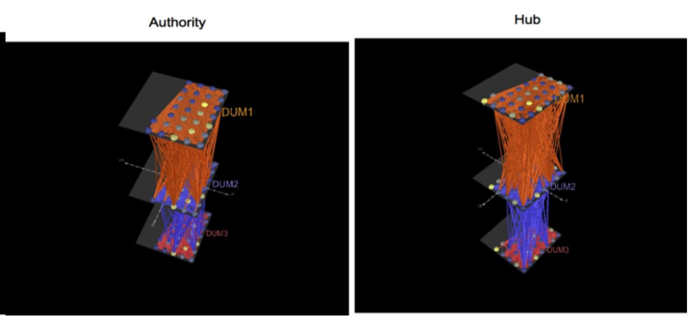


((b)) Synthetic network 1 HITS

Figure 5



((a)) Synthetic Network 2 IP



((b)) Synthetic network 2 HITS

Figure 6

The closed form solution for HITS is given by

$$h^k = \frac{(AA^T)^k * \overrightarrow{h}(0)}{||(AA^T)^k * \overrightarrow{h}(0)||_2} \tag{9}$$

$$a^k = \frac{(AA^T)^{k-1} * (A^T * \overrightarrow{h}(0)}{||(AA^T)^{k-1} * (A^T * \overrightarrow{h}(0)||_2} \tag{10}$$

Hub scores are given by dominant eigen vector of $A * A^T$.

6. CONCLUSIONS

In this project we were able to come up with a framework that analyses a node's or entity's influence in the context of multiple behaviours enacted by players of the network. We devised a model for Influence calculation leveraging the concept of Passivity and vice versa for a generalised network and developed an efficient algorithm on similar lines as the HITS algorithm. IP scores, Authority and Hub scores were calculated for multilayer networks using the same concept of IP and HITS Algorithm for single layer networks. These scores were calculated for synthetic networks as well as real world network generated with DBLP dataset. Interesting observations were made on the IP values of the various nodes and we tried to find a correlation between the single layer IP as well the inter layer connections. The observations indicate a good depiction of Influence and Passivity of nodes over the entire multilayer network using our IP Algorithm.

7. FUTURE WORK

This study can find its application in multiple domains like engineering, health care, urban transportation, social networks, etc. As most of the real world applications are aimed at certain objectives that could be modelled probabilistically, by estimating a relative distribution of the objectives among layers of multilayer network, a single network satisfying the given objective could be reduced and influence and passivity of such reduced network will produce much required inference and fine grained understanding of the roles of nodes.

Assuming a joint distribution of the objective against layers of the multilayer network, a network could be reduced by weighted summation of individual layers in the network. For example, in a transportation network, the probability distribution could be assigned by considering different objectives like the cost of transportation, congestion etc. which could be easily put to good use by metropolitan authorities in urban planning and development. A simple multilayer network's equivalent parameters like A^δ, R^δ, A^* and R^* can be reduced using the following equations described below. This proves to scale well and computationally economical when compared to some of the contemporary approaches.

$$A^\delta = \sum P_r(objective, layer_i) * A(i)$$

$$R^\delta = \sum P_r(objective, layer_i) * R(i)$$

$$A^* = \sum P_r(objective, layer_i, layer_j) * A(i, j)$$

$$R^* = \sum P_r(objective, layer_i, layer_j) * R(i, \forall k)$$

8. REFERENCES

[1] Daniel M. Romero, Wojciech Galuba, Sitaram Asur, Bernardo A. Huberman. *Influence and Passivity in Social Media*. Proceedings of the 20th international conference companion on World wide web, New York, 2011

[2] Mikko Kivelä, Alexandre Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, Mason A. Porter. *Multilayer Networks* J. Complex Netw. 2(3): 203-271 (2014)

[3] Emanuele Strano, Saray Shai, Simon Dobson, Marc Barthelemy. *Multiplex networks in metropolitan areas: generic features and local effects*

[4] Peter J. Mucha, Thomas Richardson, Kevin Macon, Mason A. Porter, Jukka-Pekka Onnela. *Community Structure in Time-Dependent, Multiscale and Multiplex Networks.*

[5] Tang, L., Wang, X. Liu, H. *Community detection via heterogeneous interaction analysis.* Data Min. Knowl. Discov., 25, 1–33.

[6] Barrett, L., Henzi, S. P. Lusseau, D. *Taking sociality seriously: the structure of multidimensional social networks as a source of information for individuals.* Phil. Trans. R. Soc. B, 367, 2108–2118.

[7] Ng, M. K.-P., Li, X. Ye, Y *MultiRank: co-ranking for objects and relations in multi-relational data.* Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'11, New York, NY, USA. ACM, pp. 1217–1225.