A centrality measure to quantify social capital in a network

Katie Allmeroth and Jordan Cowart
Department of Computer Science
Southeast Missouri State University
1 University Plaza, Cape Girardeau, MO 63701, USA
{kmallmeroth1s, jrcowart1s} @semo.edu

Abstract

As online social networks continue to become an ever pervasive part of the human social experience, the value of their study also increases. Of particular interest is the ability to determine important, or central, nodes in a network. Current measures of centrality take into account a nodes "connectedness" but fail to consider the diversity of resources available to it; its social capital. The purpose of this paper is to define social capital in relation to a network as well as describe a method by which to score it.

1. Introduction

The ability to distinguish the most important nodes within a network is a widely discussed topic and has resulted in a great number of algorithms to accomplish this task; these are known as centrality measures. With regards to social networking, however, these measures fall short. The current measures focus on analyzing the connectedness of nodes within a graph, with more connections resulting in a higher score.

In a social network, connectedness holds a great deal of relevance in deciding the importance of a node but we propose that it should not be the only factor in forming this decision. Other factors to consider include: the nodes inclusion to important communities, its connections with members of other important communities, and its access to a diversity of resources within the network. We consider the combination of the all these factors as the social capital of a node. The purpose of this paper is to give a more formal and rigid definition of this concept of social capital as well as describe a method by which to score it.

The rest of the paper is organized as follows: Section II provides an overview of the works related to this study. Section III discusses what social capital is, its relevance to social networking, and its formal definition. Section IV defines an algorithm by which social capital can be quantitatively measured. Section V discusses our methods of data collection along with the analysis and visualization of this data. Section VI summarizes our findings and provides a layout for future work.

2. Related Work

Ilyas et al. [1, 2] demonstrated the shortcomings of Eigenvector Centrality (EVC) with regards to community detection then proposed and defined a new algorithm, Principal Component Centrality (PCC). Yang et al. [3] discussed the use of ground-truth (user defined) communities as a means of benchmarking a community detection algorithm's effectiveness. Xiao et al. [4] developed a new algorithm for gathering data from Facebook by use of scrapping instead of using Facebook's API, and demonstrated its effectiveness versus the API method. The correlation between likes and community membership were discussed in [5]; using this information we will assume that an individual liking a page is a declaration of membership within that particular community.

3. A Definition of Social Capital

Kris is currently working on formalizing a method for scoring social capital based on our earlier definition. Once formalized, this section will discuss the reasoning behind this definition and its relevance to social networking. However, we do have a working definition for social capital. In this sense social capital is a measure of a nodes criticality within a graph as well as its access to a diversity of resources (e.g. its membership within multiple communities). We hope to encompass this idea within the formal definition. This biggest issue with developing this definition is determining how to weigh the connections and community membership in an effective manner.

4. A Method for Scoring Social Capital

As we do not currently have a formal definition of social capital, we do not have a formal method for scoring it. However, we do have a working definition of it, so we know what we are working towards. At this time we are examining EVC/PCC as discussed in [2] to be the key part of our algorithm. This section will have images of graphs that have been processed by an EVC/PCC algorithm to help visualize and explain their usage. The following is what has been accomplished thus far.

There are currently many methods of determining a node's measure of centrality [6]; however the two that hold the most interest in determining a node's social capital are Eigenvector Centrality (EVC) and Principal Component Centrality (PCC).

EVC is a very popular centrality measure used within the social sciences, with PCC essentially being an extension of EVC [1]. EVC quantifies centrality by recursively analyzing the weight and number of connections between nodes; this method has a notable shortcoming however. Because of the recursive nature of the algorithm and the use of only the largest eigenvector, the results are typically skewed to a particular region of the graph [1]. Since social capital should take into account a node's ability to access and influence to important resources and communities across a network, the scope needs to be expanded, hence the interest in PCC. PCC extends the idea of using eigenvectors as a key tool in determining centrality. However, instead of using solely the largest eigenvector, as EVC does, PCC uses the P largest feature vectors (eigenvectors) where P $\leq$ the total number of positive eigenvectors. This method highlights the P-largest communities instead of only using scores from the dominant eigenvector. By using the features of these algorithms, we will show how to score a nodes social capital.

5. Data Collection and Analysis

5.1 Data Collection

Our crawler is near completion but we cannot start crawling until we receive permission from Facebook.

Below is an outline of what this section will look like. Data was obtained from Facebook using a crawler based on the algorithm discussed in [4]. The algorithm was implemented in Python, utilizing Beautiful Soup (a Python library for parsing HTML) and built-in libraries. The crawler was run on a [specs of machine] for [period of time] and collected [number of nodes]. The crawler was refined in order to only scrape Facebook pages of individuals that attend the University of Missouri. We scraped the following information: friends, gender, academic major, hometown and University of Missouri related likes. All user identifiable information was altered to prevent identification. The process for crawling is seen in Algorithm 1.

Algorithm 1: Web Crawler

```
1.  Login to Facebook
2.  Initialize queue with seed node
3.  Step 1:
4.  If queue empty:
5.      Go to 23
6.  Pop node
7.  Go to about page and grab desired information
8.  Go to node like page and use modified version of algorithm in [4] to grab like information
9.  Go to node friend page and grab friends list using algorithm in [4]
10. Step 2:
11. While node friend queue not empty:
12.     Pop node
13.     Check privacy settings of new node
14.     If private:
15.         Skip node; go to Step 2
16.     Check if node attends the University of Missouri
17.     If does not attend:
18.         Add to black list; go to Step 2
19.     Add node to queue if not already inspected
20.     Insert node into parent node's friend list
21.     Go to Step 2
22. Go to Step 1
23. End
```

5.2 Data Analysis

We currently don't have any of our own data to analyze. This section will focus on the performance of our centrality algorithm and attempt to show that it can determine nodes that possess the highest amount of social capital in accordance with the definition. Diagrams and images from the results will be displayed.

6. Conclusion

This section will depend entirely upon the results of our data analysis in part V section B. Our hope is that the results will be encouraging to the further development and study of social capital in relation to social networks.

7. Acknowledgements

This work was sponsored by the National Science Foundation of the United States.

8. References

[1] M. U. Ilyas, H. Radha, "A KLT-inspired Node Centrality for Identifying Influential Neighborhoods in Graphs", 2010 44th Annual Conference on Information Sciences and Systems, IEEE, Princeton, NJ, March 2010, pp. 1-7.

[2] M. U. Ilyas, H. Radha, "Identifying Influential Nodes in Online Social Networks Using Principal Component Centrality", The 2011 IEEE International Conference on Communications , IEEE, Kyoto, June 2011, pp. 1-5.

[3] J. Yang, J. Leskovec, "Defining and Evaluating Network Communities based on Ground-truth", The 12th IEEE International Conference on Data Mining , IEEE, Brussels, Dec. 2012, pp. 745-754.

[4] Z. Xiao, B. Liu, H. Hu, T. Zhang , "Design and Implementation of Facebook Crawler Based on Interaction Simulation", The 11th IEEE International Conference On Trust, Security And Privacy In Computing And Communications , IEEE, Liverpool, June 2012, pp. 1109-1112.

[5] N. Salamos, E. Voudigari, T. Papageorgiou, M. Vazirgianni,, "Design and Implementation of Facebook Crawler Based on Interaction Simulation", 2012 IEEE International Conference on Green Computing and Communications, IEEE, Athens, Nov. 2012, pp. 368-371.

[6] J. Xie, S. Kelley, B. Szymanski, "Overlapping Community Detection in Networks: The State-of-the-Art and Comparative Study", ACM Comput. Surv. 45, 4, Article 43 (August 2013), 35 pages.

11. Copyright forms and reprint orders

You must include your fully-completed, signed IEEE copyright release form when you submit your paper. We **must** have this form before your paper can be published in the proceedings. The copyright form is available as a Word file, <copyright.doc>, as a PDF version, <copyright.pdf>, and as a text file in <authguid.txt>.

Reprints may be ordered using the form provided as <reprint.doc> or <reprint.pdf>.