Social Capital: The Power of Influencers in Networks*

(Extended Abstract)

Karthik Subbian
University of Minnesota, MN
karthik@cs.umn.edu

Dhruv Sharma
University of Minnesota, MN
dsharma@cs.umn.edu

Zhen Wen
IBM T.J. Watson Research Center, NY
zhenwen@us.ibm.com

Jaideep Srivastava
University of Minnesota, MN
srivasta@cs.umn.edu

ABSTRACT

The problem of finding the influencers in social networks has been traditionally dealt in an optimization setting by finding the top-k nodes that has the maximum information spread in the network. These methods aim to find the influencers in a network through the process of information diffusion. However, none of these approaches model the individual social value generated by collaborations in these networks. Such social value is often the real motivation for which the nodes connect to each other. In this work, we propose a framework to compute this network social value using the concept of social capital, namely the amount of bonding and bridging connections in the network. We first compute the social capital value of the network and then allocate this network value to the nodes of the network. We establish the fairness of our allocation using several axioms of fairness. Our experiments on the real data sets show that the computed social capital is an excellent proxy for finding influencers and our approach outperforms several popular baselines.

Categories and Subject Descriptors

J.4 [**Social and Behavorial Sciences**]: Sociology; I.2 [**Artificial Intelligence**]: General

General Terms

Algorithms; Experimentation; Human Factors

Keywords

Social Capital; Influencer Mining; Collaborative Networks; Information Diffusion

1. INTRODUCTION

The problem of finding influencers is becoming increasingly important. The existing methods to find influencers aim to propagate influence through the process of information diffusion. These methods infect a set of seed nodes to achieve maximum amount of information spread in the network. In these related work [7, 3], the model captures only the process of information diffusion in network and does not attempt to capture the actual social value of the person in that network. The notion of the influence model used in these papers fails to offer insights on the influence in the context of whole network, or when there are few available observation of information flow. To explain, let us consider the following example. The new CEO of a company may have only a few connections and limited information flows in the social networks within the organization. Whether he/she can influence a new technology in the company? The answer would be yes. The reason for this influence is not because of his/her few connections or limited information flows, but because of the control that he/she exerts on the network resources (in this case all the employees of the company). Such aspects, cannot be fully captured by local interactions of two nodes. Instead, they can only be studied if we understand the value that each node contributes to and derive from the overall network. We hypothesize that *nodes with high social value in networks tend to be more influential*, as the case of the CEO in our example.

For this purpose, we first characterize the value of the network using "Social Capital". There are various definitions of social capital [2, 4]. One of the most accepted and general definition is: *"Social capital is about the value of social networks, bonding similar people and bridging diverse people, with norms of reciprocity"* [4]. The **bonding capital** is the ability to calibrate people against each other, and **bridging capital** is the ability to connect diverse sets of people. The power of these bonding and bridging nodes to cooperate and communicate with each other creates an inherent value for the entire network. The overall value generated by such cooperation is termed as the *social capital of the network*. Now the question remaining is, how the nodes share this overall social capital amongst themselves in a fair way? After we know the fair share of network value for each node, we hypothesize that this value is proportional to the potential of a node to influence the network.

2. OUR APPROACH

We first begin by formalizing the notion of social capital value $v(g)$ for the network g. For this purpose, we follow the definition in [4]: "social capital is the value of social networks, bonding similar people and bridging between diverse people, with norms of reciprocity". Given a network $g = \langle V, E \rangle$ with vertex set V and edge set E. We define the social capital value $v(g)$ as, $v(g) = \sum_{(i,j)\in E} b\left(d_g(i,j)\right)$. The distance function d_g computes the length of the shortest path between nodes i and j. Note that, we as-

*This work was supported by the DARPA grant number W911NF-12-C-0028.

Appears in: *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013)*, Ito, Jonker, Gini, and Shehory (eds.), May, 6–10, 2013, Saint Paul, Minnesota, USA.

sume that people often make new connections within the network to reach newer friends through shortest paths. This assumption is consistent with many network formation studies [6, 8]. The function $b(.)$ is the benefits acheived due to shortest path of certain length l. We choose the benefits function to be exponentially decaying $e^{-\lambda l}$, where l is the length of the shortest path. When there is no shortest path between i and j the function d_g tends to ∞ and the corresponding benefit value is zero.

This definition captures the two important desired aspects of our valuation function: (1) the benefits from immediate neighbors, and (2) the benefits from non-immediate neighbors which decay as a function of social distance. In addition, it also satisfies two important valuation properties, anonymity and component additivity [8]. The anonymity property ensures that any permutation of node labels will not alter the node valuations. The property of component additivity ensures that the total value of the disconnected components of the network g is the sum of the individual components, as there is no surplus generated by any cooperation between these components.

We define *allocation function* $Y : \Psi \rightarrow \Re^n$, where Ψ is the set of all possible network topologies of node set V and $Y = [Y_1 ... Y_n]$ denotes the allocated social capital value of nodes 1 through n. Our allocation function satisfies the following four desired properties Y: (1) anonymity; (2) component balance; (3) weak link symmetry; and (4) improvement property. Therefore, it falls in the same class as the *Myerson Value* allocation function [5].

The anonymity property ensures that the allocation function is independent of the player labels. The property of weak link symmetry is a more general form of equality criterion specified in the Myerson's allocation function. This states that when a new edge e is added to the network g, if the utility of one end point increase then the other node's utility must strictly increase. We prefer to ensure this criterion compared to the equality criterion, because the utility received by adding a new edge in the network may not be necessarily due to equal contributions from both nodes. Our criteria guarantees that when adding a new edge $e = (i, j)$ to the network g, if the utility for any node other than i and j increases, then the utility for at least one of the nodes i or j must strictly increase.

Our proposed allocation function Y is based on the idea that each node contributes a certain value to the network by being in the shortest path of lengths l varying from 1 to $|V - 1|$. We measure their fractional contribution of each node $k \in V$ for all possible shortest path length and take a weighted average of the benefits due to each path length based on this fractional contribution. Let this fractional contribution for path length l for node k be denoted as $\alpha_k^l(g)$. The allocation function $Y_k(g)$ for any node $k \in V$ is defined as the weighted sum of benefits due to all the path lengths, weighted by the corresponding fractional contribution. More formally, $Y_i(g) = \sum_{l=1}^{n-1} \alpha_i^l(g) b(l)$.

Algorithm: Our algorithm to implement the allocation rule is listed in Algo. 1. The core part of the algorithm $SCVal$ is the sub algorithm $ComputeFC$, which computes α_k^l, the fractional contribution of node k for paths of length l. Once α_k^l gets computed for all $k \in V$ and for all path lengths l varying from 1 to L, the algorithm $SCVal$ simply sums up the weighted fractional contributions for each node k in Y_k, where the benefit function $b(l)$ acts as the corresponding weight for each path length l. Finally, vector Y is returned as the output by $SCVal$. The parameter L can be adjusted to control the maximum path length l in the algorithm.

The algorithm for computing the all pair shortest paths has the time complexity of $O(|V|^3)$, which is computationally infeasible for the large networks. Hence we developed $ComputeFC$ as a

Algorithm 1: SCVal

Input: g: social network graph; L: maximum path length
Output: $Y = [Y_1, ..., Y_n]$: social capital of nodes in V

$Y \leftarrow 0$;
$\alpha_v^l \leftarrow ComputeFC(g, L)$;
for $i \leftarrow$ *1* **to** n **do**
  for $l \leftarrow$ *1* **to** L **do**
    $Y_i \leftarrow Y_i + \alpha_i^l * b(l)$;
Return Y;

modified version of Brandes betweenness centrality algorithm [1]. The complexity of Brandes algorithm is $O(|V||E|)$, which is a significant improvement over $O(|V|^3)$ for sparse networks. The Brandes algorithm [1], however, *does not compute the fractional contributions* based on varying path lengths. Our proposed algorithm $ComputeFC$ calculates the fractional contributions for varying path lengths for each node k at *no additional cost*.

We evaluated our algorithm with real-life collaboration networks such as DBLP [1] and USPTO [2] against the popular baselines such as PMIA [3], PageRank, and weighted degree. We find that our algorithm outperforms the baseline algorithms in terms of the standard influence measure, such as the expected number of infected nodes [7]. The run times of our algorithm was better than PMIA and comparable to that of the other two baselines.

3. CONCLUSION

In summary, we proposed a new approach to compute influencers in networks using their social capital value. We formulated the problem of computing social capital and sharing this value amongst the nodes of the network as a value-allocation model that satisfies several desired properties. We propose an efficient algorithm and empirically demonstrate the effectiveness using two real-life data sets.

4. REFERENCES

[1] U. Brandes. A faster algorithm for betweenness centrality. In *Jour. of Math. Soc.*, volume 25, pages 163–177, 2001.

[2] R. S. Burt. *Brokerage and Closure : An Introduction to Social Capital*. Oxford University Press, 2005.

[3] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *ACM KDD*, pages 1029–1038, 2010.

[4] P. Dekker and E. M. Uslaner. *Social Capital and Participation in Everyday Life*. Rouledge, 2001.

[5] B. Dutta, A. V. D. Nouweland, and S. Tijs. Link formation in cooperative situtations. In *Int. Jour. of Game Theo.*, volume 27, pages 245–256, 1998.

[6] M. O. Jackson and A. Wolinsky. A strategic model of social and economic networks. In *Jour. of Eco. Theo.*, volume 71, pages 44–74, 1996.

[7] D. Kempe, J. M. Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *ACM KDD*, pages 137–146, 2003.

[8] R. Narayanam and Y. Narahari. Topologies of strategically formed social networks based on a generic value function - allocation rule model. In *Social Networks*, volume 33, pages 56–69, 2011.

[1] http://dblp.uni-trier.de/xml/

[2] http://www.google.com/googlebooks/uspto.html