

Identification of Highly Influential Individuals in a Rural Social Network

Nandini Bhardwaj
Digital Green
New Delhi, India
nandini@digitalgreen.org

Yash Seth
Digital Green
New Delhi, India
yash@digitalgreen.org

ABSTRACT

The purpose of this paper is to gain insights to improve effectiveness of an intervention. This is achieved by visualizing the flow of information through a social network constructed in order to motivate behavior change. Early adopters in this social network are identified by computing an F-score[2] for individuals, which corresponds to their influence within their communities. This influence decays with distance across the network. Thus, highly influential individuals in communities are identified. This information can potentially be used to improve the efficiency of this intervention.

Categories and Subject Descriptors

H.3.7 [Information Storage and Retrieval]: Digital Libraries—Dissemination; G.m [Probability and Statistics]: Miscellaneous; E.1 [Data Structures]: Graphs and networks

1. INTRODUCTION

The goal of many social development interventions is to transform communities by altering behavior patterns. In an ideal situation, merely introducing individuals in a community to beneficial behavior would result in immediate adoption. However, the social context within which a beneficiary exists remains a powerful influence on personal choices and is much more resistant to change. Thus, the effectiveness of interventions which target a lasting change in society through raising awareness and disbursement of information can be measured through adoption of new behaviors and transformation in repetitive behaviors. Some interventions seek to impose an independent social structure on top of, or in conjunction with, existing networks to facilitate information flow. In this paper, data collected through the Digital Green intervention is analyzed. This intervention functions through a network of grassroots workers who support beneficiaries in adopting new agricultural practices. Beneficiaries are organized as Self-Help Groups to facilitate dialogue and mutual support. At higher levels, support is provided by professionals who

facilitate coordination within a geographical entity, such as a district. This constructed social structure can be seen as a network through which information is being diffused[3]. This paper seeks to devise a way to improve the effectiveness of such an intervention. We will do so visualizing the flow of information through this network by employing statistical and graphical means to identify influential individuals which drive the most change.

2. METHODOLOGY

2.1 Description of the Network

Several different network topologies[1] are possible in the social processes being represented. The topology chosen here is a hierarchical social network depicting information flow with different levels representing geographical units and sub-units, such as village, block and district. In this network, nodes represent individuals and edges represent social relationships they have as a result of participation in the intervention. Leaf nodes represent beneficiaries who participate in the intervention and may or may not adopt behaviors demonstrated and advocated.

2.2 Network Hierarchy

In the larger context of the intervention, there are five levels of hierarchy. At the top of the hierarchy is a state coordinator whose role involves facilitating planning and coordination between different districts. The district coordinator has a similar role at the district level and facilitates content sharing and coordination between block officers. The block officer coordinates interactions and knowledge sharing and support between village resource persons. The specific subset of data that is chosen for the purpose of this paper represents a single district and has three levels of hierarchy.

2.3 Information Flow

A node goes through three different states during information flow. At the beginning, a node does not have access to the information at all. Once it gets the information, it becomes open to influence. At this point, it may choose to act on the information and adopt it. If it adopts the information, it becomes a source of influence. Nodes which are highly influential are highly predictive of future adoptions by other nodes within their communities.

In order to identify influential nodes, we compute the number of adoptions which took place after the node in contention becomes a source of influence; in our metric of choice, the F-score[4], this is the number of true positives.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DEV '13, January 11-12, 2013 Bangalore India
Copyright 2013 ACM 978-1-4503-1856-3/13/01 ...\$15.00.

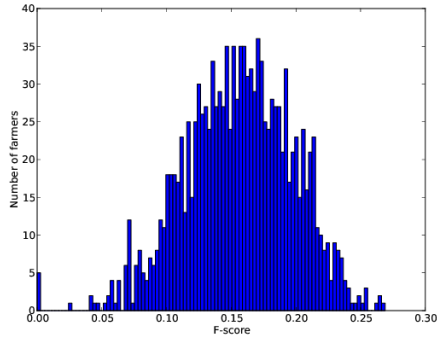


Figure 1: Distribution of F-scores across farmers in district Guna

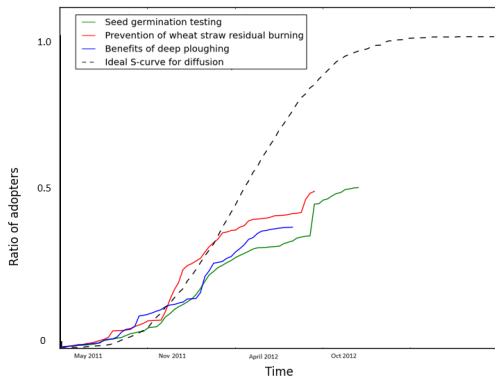


Figure 2: Diffusion patterns of three popular videos in district Guna

Next, we count the number of nodes which did not adopt the practice or adopted the practice before the node became a source of influence and use this as the number of predictions of adoptions which did not materialize, i.e. false positives. Similarly, we compute true negatives, i.e. the number of individuals who did not adopt the practice given that the source node did not adopt, and false negatives, i.e. the number of individuals who adopt the practice given that the source node did not adopt. Combining these four values, we compute a confusion matrix for the predictive power of each individual node. We use this confusion matrix to compute an F-score, which acts as an indicator of the influencing capacity of a source node and thereby helps in identifying influencers. Figure 1 shows a histogram of F-score values obtained through the above computation.

3. DATASET

In this study, we have employed the approach described above on a dataset detailing the social intervention efforts of Digital Green and ACCESS Development Services in Guna district of Madhya Pradesh. This dataset describes the activities of 1280 farmers who have adopted 133 distinct practices between June 2010 and September 2012. Information has been diffused through this network through 7480 screenings and there have been 13622 instances of adoption of new behaviors in these communities.

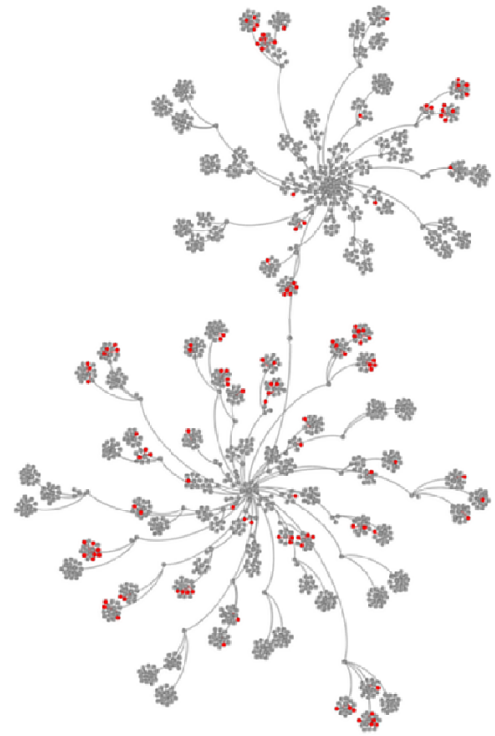


Figure 3: Network of individuals in Guna district. Red nodes depict top influencers.

4. RESULTS AND CONCLUSIONS

Diffusion patterns of three videos in Guna district can be seen in Figure 2. It shows the adoption pattern of these videos along with the ideal S-curve for diffusion. The F-score metric has been used to identify the top 10% of farmers in terms of influence. Figure 3 is a visualization of this network, with influential individuals marked in red. One such influential farmer identified by this algorithm is Nilam who has been associated since September 2010. She is well respected in her community, and has even been featured in a video produced for dissemination. Our preliminary results indicate that this method is effective in identifying highly influential individuals. In the future, we hope to demonstrate the potential for increasing effectiveness in information dissemination through these individuals.

5. REFERENCES

- [1] M. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.
- [2] C. J. V. Rijsbergen. *Information Retrieval*. Butterworth-Heinemann, 1979.
- [3] E. M. Rogers. *Diffusion of innovations (5. ed.)*. Free Press, 2003.
- [4] S. V. Stehman. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62(1):77 – 89, 1997.