HHS Public Access

Author manuscript

Soc Netw. Author manuscript; available in PMC 2015 March 31.

Published in final edited form as:

Soc Netw. 2015 January 1; 4(1): 1–16. doi:10.4236/sn.2015.41001.

Estimating Vertex Measures in Social Networks by Sampling Completions of RDS Trees

Bilal Khan[1], **Kirk Dombrowski**[2], **Ric Curtis**[2], and **Travis Wendel**[3]

[1]Department of Math and Computer Science, John Jay College (CUNY), New York, USA

[2]Department of Sociology, University of Nebraska-Lincoln, Lincoln, USA

[3]St. Ann's Corner of Harm Reduction, Bronx, USA

Abstract

This paper presents a new method for obtaining network properties from incomplete data sets. Problems associated with missing data represent well-known stumbling blocks in Social Network Analysis. The method of "estimating connectivity from spanning tree completions" (ECSTC) is specifically designed to address situations where only spanning tree(s) of a network are known, such as those obtained through respondent driven sampling (RDS). Using repeated random completions derived from degree information, this method forgoes the usual step of trying to obtain final edge or vertex rosters, and instead aims to estimate network-centric properties of vertices probabilistically from the spanning trees themselves. In this paper, we discuss the problem of missing data and describe the protocols of our completion method, and finally the results of an experiment where ECSTC was used to estimate graph dependent vertex properties from spanning trees sampled from a graph whose characteristics were known ahead of time. The results show that ECSTC methods hold more promise for obtaining network-centric properties of individuals from a limited set of data than researchers may have previously assumed. Such an approach represents a break with past strategies of working with missing data which have mainly sought means to complete the graph, rather than ECSTC's approach, which is to estimate network properties themselves without deciding on the final edge set.

Keywords

Network Imputation; Missing Data; Spanning Tree Completions; Respondent-Driven Sampling

1. Introduction

Respondent-Driven Sampling (RDS) has become a popular technique for providing statistically meaningful data on hard to reach populations by using peer-referral methods.

bkhan@jjay.cuny.edu, kdombrowski2@unl.edu, rcurtis@jjay.cuny.edu, twendel@sachr.org

How to cite this paper: Khan, B., Dombrowski, K., Curtis, R. and Wendel, T. (2015) Estimating Vertex Measures in Social Networks by Sampling Completions of RDS Trees. *Social Networking*, 4, 1-16. http://dx.doi.org/10.4236/sn.2015.41001

Table 2

Misclassification (mean and standard deviation) over 25 trials.

Measure: BC				
$\bar{r}$	**1 comps**	**10 comps**	**30 comps**	**50 comps**
1 trees	11.404	9.596	9.762	9.895
10 trees	9.814	11.088	11.389	11.561
30 trees	10.667	11.596	11.812	11.784
50 trees	10.869	11.735	11.895	11.868

std ε	**1 comps**	**10 comps**	**30 comps**	**50 comps**
1 trees	1.035	0.641	0.589	0.592
10 trees	0.476	0.414	0.281	0.264
30 trees	0.439	0.271	0.176	0.167
50 trees	0.462	0.233	0.161	0.174

Measure: ES				
$\bar{r}$	**1 comps**	**10 comps**	**30 comps**	**50 comps**
1 trees	8.447	7.872	7.842	7.843
10 trees	7.862	7.838	7.838	7.838
30 trees	7.839	7.838	7.838	7.838
50 trees	7.838	7.838	7.838	7.838

std ε	**1 comps**	**10 comps**	**30 comps**	**50 comps**
1 trees	0.460	0.070	0.034	0.036
10 trees	0.051	0.000	0.000	0.000
30 trees	0.003	0.000	0.000	0.000
50 trees	0.000	0.000	0.000	0.000

Measure: CON				
$\bar{r}$	**1 comps**	**10 comps**	**30 comps**	**50 comps**
1 trees	13.836	11.617	11.521	11.584
10 trees	11.652	11.593	11.579	11.578
30 trees	11.550	11.578	11.575	11.575
50 trees	11.598	11.575	11.575	11.575

std ε	**1 comps**	**10 comps**	**30 comps**	**50 comps**
1 trees	1.085	0.303	0.190	0.126
10 trees	0.358	0.020	0.009	0.008
30 trees	0.112	0.006	0.000	0.000
50 trees	0.043	0.000	0.000	0.000